



GARY GROTH-MARNAT
A. JORDAN WRIGHT

HANDBOOK OF
**PSYCHOLOGICAL
ASSESSMENT**

SIXTH EDITION

WILEY

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

Sixth Edition

Gary Groth-Marnat and A. Jordan Wright

WILEY

Copyright © 2016 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, or on the Web at www.copyright.com. Requests to the publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 201-748-6011, fax 201-748-6008, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Readers should be aware that Internet Web sites offered as citations and/or sources for further information may have changed or disappeared between the time this was written and when it is read.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If legal, accounting, medical, psychological or any other expert assistance is required, the services of a competent professional should be sought.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 800-956-7739, outside the U.S. at 317-572-3986, or fax 317-572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Names: Groth-Marnat, Gary, author. | Wright, A. Jordan, author.

Title: Handbook of psychological assessment / Gary Groth-Marnat and A. Jordan Wright.

Description: Sixth edition. | Hoboken, New Jersey : John Wiley & Sons, Inc., [2016] | Includes index.

Identifiers: LCCN 2015042889 (print) | LCCN 2016000499 (ebook) | ISBN 9781118960646 (cloth) | ISBN 9781118960684 (pdf) | ISBN 9781118960653 (epub)

Subjects: LCSH: Psychological tests. | Personality assessment.

Classification: LCC BF176 .G76 2016 (print) | LCC BF176 (ebook) | DDC 150.28/7—dc23

LC record available at <http://lcn.loc.gov/2015042889>

Cover Image: © 501room/Shutterstock

Cover Design: Wiley

Printed in the United States of America

SIXTH EDITION

HB Printing 10 9 8 7 6 5 4 3 2 1

Contents

Preface	xi
Chapter 1 Introduction	1
Organization of the Handbook	1
Role of the Clinician	3
Patterns of Test Usage in Clinical Assessment	6
Evaluating Psychological Tests	10
Validity in Clinical Practice	23
Clinical Judgment	26
Phases in Clinical Assessment	32
Recommended Reading	37
Chapter 2 Context of Clinical Assessment	39
Types of Referral Settings	39
Ethical Practice of Assessment	50
Assessing Diverse Groups	59
Selecting Psychological Tests	67
Computer-Assisted Assessment	72
Recommended Reading	74
Chapter 3 The Assessment Interview	77
History and Development	78
Issues Related to Reliability and Validity	82
Assets and Limitations	84
The Assessment Interview and Case History	86
Mental Status Evaluation	93
Interpreting Interview Data	99
Structured Interviews	100
Recommended Reading	111

Chapter 4	Behavioral Assessment	113
	History and Development	115
	Issues Related to Reliability and Validity	117
	Assets and Limitations	120
	Strategies of Behavioral Assessment	121
	Recommended Reading	136
Chapter 5	Wechsler Intelligence Scales	139
	Testing of Intelligence: Pros and Cons	139
	History and Development	142
	Reliability and Validity	147
	Assets and Limitations	151
	Use with Diverse Groups	153
	Meaning of IQ Scores	156
	Cautions and Guidelines in Administration	159
	WAIS-IV/WISC-V Successive-Level Interpretation Procedure	161
	Wechsler Indexes and Subtests	173
	Assessing Brain Damage	196
	Assessing Additional Special Populations	204
	Short Forms	209
	Recommended Reading	213
Chapter 6	Wechsler Memory Scales	215
	History and Development	216
	Reliability and Validity	222
	Assets and Limitations	224
	Use with Diverse Groups	226
	Interpretation Procedure	227
	Interpreting Patterns of Index Scores	230
	Comparing Scores on the WAIS-IV and the WMS-IV	238
	Additional Considerations: Malingering and Evaluating Change	241
	Recommended Reading	242
Chapter 7	Minnesota Multiphasic Personality Inventory	243
	History and Development	246
	Reliability and Validity	251
	Assets and Limitations	255
	Use with Diverse Groups	259

Administration	262
MMPI-2 Interpretation Procedure	263
MMPI-2 Computerized Interpretation	272
MMPI-2 Validity Scales	273
MMPI-2 Clinical Scales	279
MMPI-2 2-Point Codes	302
MMPI-2 Content Scales	328
MMPI-A Content Scales	331
MMPI-2 Harris-Lingoes and <i>Si</i> Subscales	334
MMPI-2 Critical Items	337
MMPI-2 and MMPI-A Supplementary Scales	338
MMPI-2-RF Interpretation Procedure	341
MMPI-2-RF Validity Scales	346
MMPI-2-RF Higher-Order Scales	353
MMPI-2-RF Restructured Clinical Scales	355
MMPI-2-RF Specific Problem Scales	361
Interest Scales	366
MMPI-2-RF Personality Psychopathology Five Scales	366
Recommended Reading	370
Chapter 8 Personality Assessment Inventory	371
History and Development	373
Reliability and Validity	373
Assets and Limitations	375
Use with Diverse Groups	377
Interpretation Procedure	377
Validity Scales	380
Clinical Scales	384
Treatment Scales	404
Interpersonal Scales	408
Additional Clusters	410
Critical Items	415
Recommended Reading	416
Chapter 9 Millon Clinical Multiaxial Inventory	417
History and Development	419
Reliability and Validity	424
Assets and Limitations	427

Use with Diverse Groups	431
Interpretation Procedure	432
Validity Scales	436
Clinical Personality Patterns	439
Severe Personality Pathology	476
Clinical Syndromes	486
Severe Syndromes	488
Recommended Reading	488
Chapter 10 NEO Personality Inventory	489
History and Development	490
Reliability and Validity	491
Assets and Limitations	493
Use with Diverse Groups	494
Interpretation Procedure	495
Recommended Reading	512
Chapter 11 The Rorschach	513
History and Development	514
Reliability and Validity	518
Assets and Limitations	523
Use with Diverse Groups	527
Comprehensive System: Administration	528
Comprehensive System: Coding	531
Comprehensive System: Scoring the Structural Summary	536
Comprehensive System: Interpretation	539
R-PAS: Administration	581
R-PAS: Coding	584
R-PAS: Scoring the Structural Summary	592
R-PAS: Interpretation	593
Recommended Reading	614
Chapter 12 Screening for Neuropsychological Impairment	615
History and Development	617
Interviewing for Brain Impairment	620
Domains of Neuropsychological Functioning	624
Bender Visual-Motor Gestalt Test, Second Edition	633

Repeatability Battery for the Assessment of Neuropsychological Status Update	645
Recommended Reading	652
Chapter 13 Brief Instruments for Treatment Planning, Monitoring, and Outcome Assessment	653
Selecting Brief Instruments	654
Symptom Checklist-90-R and Brief Symptom Inventory	655
Beck Depression Inventory–II	662
State Trait Anxiety Inventory	666
Recommended Reading	670
Chapter 14 Treatment Planning and Clinical Decision Making	671
Development and Approaches to Treatment Planning	672
Intervention Options	677
Clinical Decision Making	679
Case Formulation	680
Understanding the Problems	683
Understanding Problem Context	692
Treatment-Specific Client Characteristics	696
The Systematic Treatment Selection (STS)/Innerlife Approach	703
Recommended Reading	705
Chapter 15 The Psychological Report	707
General Guidelines	708
Feedback	728
Format for a Psychological Report	730
Sample Reports	743
Recommended Reading	772
Appendix A Test Publishers/Distributors	773
Appendix B Testing Organizations	777
References	779
Author Index	857
Subject Index	895

Preface

Thank you so much for your support in buying and reading this book. Our intention has been to create a resource that will cover the A to Z of assessment. In other words, our aim has been to provide guidance that includes larger issues on assessment as well as specific stages in the assessment process, from clarifying the referral question through writing up the report and providing feedback and consulting with your referral sources and clients. We hope it brings clarity, practical guidelines, insights, and useful strategies to your work. Feedback on the previous editions assures us that this is often the case. This fact makes it worth the many long hours hidden away inside a small room incubating ideas and reading, writing, revising, and editing.

As with the previous editions, we have tried to integrate the best of science with the best of practice. Necessarily, psychological assessment involves technical knowledge. But in presenting this technical knowledge, we have tried to isolate, extract, and summarize in as clear a manner as possible the core information that is required for practitioners to function competently. At the same time, assessment is also about the very human side of understanding, helping, and making decisions about people. We hope we have been able to comfortably blend this technical (science) side with the human. An assessment that does not have at least some heart to it is cold and falls short in understanding the experience of the client. To keep in touch with the practitioner/human side of assessment, we have continually maintained active assessment practices in which we have tried to stay close to and interact with the ongoing personal and professional challenges of practitioners. We hope that within and between the sentences in the book, our active involvement with the world of practice is apparent.

It has been seven years since the previous (fifth) edition was published. During that time, much has changed but much has remained the same. The big tests that professional psychologists use most frequently are somewhat different, and this is reflected in changes to this sixth edition. This includes eliminating the chapter on the California Psychological Inventory and replacing it with the more widely used NEO Personality Inventory—3. While both focus on normal personality traits, the NEO is based on the strongly empirically supported five-factor model of personality. Additionally, this edition has eliminated the chapter on the Thematic Apperception Test. Although its use in clinical practice is unclear (anecdotally it seems still to be relatively widely used), the test itself has suffered from the lack of consensus on a coding and scoring protocol and a subsequent lack of strong, consistent empirical support. In place of this chapter, we have included a chapter on the Personality Assessment Inventory, which has gained both strong empirical support and wide clinical popularity.

In addition to these major changes in tests covered, there are important changes within other chapters. The chapter on the Wechsler Intelligence Scales includes

updated information on the newly developed Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V). Additionally, two chapters have been significantly expanded because each test has two alternate forms, both of which currently are in wide use. Specifically, the chapter on the Minnesota Multiphasic Personality Inventory includes information on both the MMPI-2 and the MMPI-2–Restructured Form (MMPI-2-RF), and the chapter on the Rorschach includes information on both the Comprehensive System and the Rorschach Performance Assessment System (R-PAS). For both of these tests, the next 10 or so years may see the field lean toward one or the other form of the tests, but at the moment the field is split, and both versions of both tests are widely used. Finally, we have worked to strengthen the sections on “Use with Diverse Groups,” which reflect the more extensive use of assessment for a wide variety of populations and the importance of competently and sensitively working with diverse populations.

There are also many smaller changes throughout this sixth edition. It has been fully updated with new research in the field. There has also been greater emphasis on making assessment more user friendly and consumer oriented. This is reflected in suggestions for using everyday language in reports, connecting interpretations to actual client behavior, strategies for wording interpretations in a manner likely to enhance client growth, and the importance of collaborating with clients. The treatment planning and clinical decision making chapter has been completely updated, and the psychological report writing chapter has been updated to include current thinking of the American Psychological Association and the Society for Personality Assessment about proficiency in personality assessment. We hope that these changes will provide readers with the best, most current, and most practical of what can be available in assessment.

The development of the *Handbook of Psychological Assessment* has been a group effort. It started many years ago with ideas and cowriting with Gary Groth-Marnat’s colleague Dorothy Morena. We wanted to develop a resource that would assist students with all phases of psychological assessment. Our sincere thanks to her. A series of editors at John Wiley & Sons have been invaluable, including Herb Reich, Jennifer Simon, Tracey Belmont, Lisa Gebo, Peggy Alexander, and Marquita Flemming. We have very much enjoyed and appreciated our relationship with Wiley; not only have we been treated as respected authors, but they have also welcomed us into the Wiley “family.” Colleagues who have provided valuable input include Steve Smith, Larry Beutler, Steve Finn, Alan Kaufman, Dawn Flanagan, Greg Meyer, Joni Mihura, Aasha Foster, and the invaluable and nonstop list of articles from the Kenneth Pope website and listerv. Seth Grossman, C. J. Thompson, and their colleagues at Pearson Assessment were extremely helpful and generous in supplying us with advance information on the MCMI-IV. Finally, much of our professional work is devoted toward helping students to achieve the best of what they are capable of. In return, working with them has inevitably helped us refine this sixth edition. Finally, we would like to dedicate the sixth edition to Gary’s parents, Barbara and Rudy, in memoriam, as well as to Jordan’s husband, Matt, and daughter, Millie, for their unwavering support.

Gary Groth-Marnat and A. Jordan Wright
July 28, 2015

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

INTRODUCTION

The *Handbook of Psychological Assessment* is designed to develop a high level of practitioner competence by providing relevant, practical research, and theoretical information. It can serve as both a reference and an instructional guide. As a reference book, it aids in test selection and the development of a large number and variety of interpretive hypotheses. As an instructional text, it provides students with the basic tools for conducting an integrated psychological assessment. The significant and overriding emphasis in this book is on assessing areas that are of practical use in evaluating individuals in a clinical context. It is applied in its orientation, and, for the most part, theoretical discussion has been kept to a minimum. Many books written on psychological testing and the courses organized around these books focus primarily on test theory, with a brief overview of a large number of tests. In contrast, the intent of this book is to focus on the actual processes that practitioners go through during assessment. We begin with such issues as role clarification and evaluation of the referral question and end with treatment planning and the actual preparation of the report itself.

One of the crucial skills that we hope readers of this text will develop, or at least have enhanced, is a realistic appreciation of the assets and limitations of assessment. This includes an appraisal of psychological assessment as a general strategy as well as an awareness of the assets and limitations of specific instruments and procedures. A primary limitation of assessment lies in the incorrect handling of the data, which are not integrated in the context of other sources of information (behavioral observations, history, other test scores). Also, the results are not presented in a way that helps solve the unique problems clients or referral sources are confronting. To counter these limitations, the text continually provides practitioners with guidelines for integrating and presenting the data in as useful a manner as possible. The text is thus not only a book on test interpretation (although this is an important component) but on test integration within the wider context of assessment. As a result, psychologists should be able to create reports that are accurate, effective, concise, and highly valued by the persons who receive them.

ORGANIZATION OF THE HANDBOOK

The central organizational plan for the *Handbook of Psychological Assessment* replicates the sequence practitioners follow when performing an evaluation. They are initially concerned with clarifying their roles, ensuring that they understand all the implications of the referral question, deciding which procedures would be most appropriate for the assessment, and reminding themselves of the potential problems

associated with clinical judgment (this chapter). They also need to understand the context in which they will conduct the assessment. This understanding includes appreciating the issues, concerns, terminology, and likely roles of the persons from these contexts. Practitioners also must follow clear ethical guidelines, know how to work with persons from diverse backgrounds, and recognize issues related to computer-assisted assessment and the ways that the preceding factors might influence their selection of procedures (see Chapter 2).

Once practitioners have fully understood the preliminary issues discussed in this chapter and Chapter 2, they must select different strategies of assessment. The three major strategies are interviewing, observing behavior, and psychological testing. An interview is likely to occur during the initial phases of assessment and is also essential in interpreting test scores and understanding behavioral observations (see Chapter 3). The assessment of actual behaviors might also be undertaken (see Chapter 4). Behavioral assessment might be either an end in itself or an adjunct to testing. It might involve a variety of strategies, such as the measurement of overt behaviors, cognitions, alterations in physiology, or relevant measures from self-report inventories.

The middle part of the book (Chapters 5 through 13) provides a general overview of the most frequently used tests. Each chapter begins with an introduction to the test in the form of a discussion of its history and development, current evaluation, and procedures for administration, as well as use with diverse populations. The main portions of these chapters provide a guide for interpretation, which includes such areas as the meaning of different scales, significant relations between scales, frequent trends, and the meaning of unusually high or low scores. When appropriate, there are additional subsections. For example, Chapter 5, “Wechsler Intelligence Scales,” includes additional sections on the meaning of IQ scores, estimating premorbid IQ, and assessing special populations. Likewise, several chapters include alternative procedures for using the tests, such as Chapter 7, “Minnesota Multiphasic Personality Inventory,” which includes procedures for both the MMPI-2 and the MMPI-2-RF, and Chapter 11, “The Rorschach,” which includes both the Comprehensive System and the R-PAS versions of the Rorschach. Chapter 12, “Screening for Neuropsychological Impairment,” varies somewhat from the preceding format in that it is more a compendium and interpretive guide to some of the most frequently used short neuropsychological tests. It also includes a section on special considerations in conducting a neuropsychological interview. This organization reflects the current emphasis on and strategies for assessing patients with possible neuropsychological dysfunction.

Several of the chapters on psychological tests are quite long, particularly those for the Wechsler intelligence scales, the Minnesota Multiphasic Personality Inventory, and the Rorschach. These chapters include extensive summaries of a wide variety of interpretive hypotheses intended for reference purposes when practitioners must generate interpretive hypotheses based on specific test scores. To gain initial familiarity with the tests, we recommend that practitioners or students carefully read the initial sections (history and development, psychometric properties, etc.) and then skim through the interpretation sections more quickly. Doing this provides the reader with a basic familiarity with the procedures and types of data obtainable from the tests. As practical test work progresses, clinicians can then study the interpretive hypotheses in greater depth and gradually develop more extensive knowledge of the scales and their interpretation.

Based primarily on current frequency of use, these tests are covered in this text: Wechsler intelligence scales (WAIS-IV/WISC-V), Wechsler Memory Scales (WMS-IV), Minnesota Multiphasic Personality Inventory (MMPI-2 and MMPI-2-RF), Millon Clinical Multiaxial Inventory (MCMI-IV), Personality Assessment Inventory (PAI), NEO Personality Inventory–3 (NEO-PI-3), Bender Visual Motor Gestalt Test–II, Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), and the Rorschach (Comprehensive System and R-PAS; Camara, Nathan, & Puente, 2000; C. Piotrowski & Zalewski, 1993; Rabin, Barr, & Burton, 2005; Watkins, 1991; Watkins, Campbell, Nieberding, & Hallmark, 1995). The NEO-PI-3 was selected because of the importance of including a broad-based inventory of normal functioning, along with its excellent technical development and relatively large research base. We have also included Chapter 13 focusing on the most frequently used brief, symptom-focused inventories because of the increasing importance of monitoring treatment progress and outcome in a cost- and time-efficient managed care environment (Eisman et al., 2000; C. Piotrowski, 1999). The preceding instruments represent the core assessment devices used by most practitioners.

Finally, the clinician must generate relevant treatment recommendations and integrate the assessment results into a psychological report. Chapter 14 provides a systematic approach for working with assessment results to develop practical, evidence-based treatment recommendations. Chapter 15 presents guidelines for report writing, a report format, and four sample reports representative of the four most common types of referral settings: medical setting, legal context, educational context, and psychological clinic. Thus, the chapters follow a logical sequence and provide useful, concise, and practical knowledge.

ROLE OF THE CLINICIAN

The central role of clinicians conducting assessments should be to answer specific questions and make clear, specific, and reasonable recommendations to help improve functioning. To fulfill this role, clinicians must integrate a wide range of data and bring into focus diverse areas of knowledge. Thus, they are not merely administering and scoring tests. A useful distinction to highlight this point is the contrast between a psychometrist and a clinician conducting psychological assessment (Maloney & Ward, 1976; Matarazzo, 1990). Psychometrists tend to use tests merely to obtain data, and their task is often perceived as emphasizing the clerical and technical aspects of testing. Their approach is primarily data oriented, and the end product is often a series of traits or ability descriptions. These descriptions are typically unrelated to the person's overall context and do not address unique problems the person may be facing. In contrast, psychological assessment attempts to evaluate an individual in a problem situation so that the information derived from the assessment can somehow help with the problem. Tests are only one method of gathering data, and the test scores are not end products but merely means of generating hypotheses. Psychological assessment, then, places data in a wide perspective, with its focus being problem solving and decision making.

The distinction between psychometric testing and psychological assessment can be better understood and the ideal role of the clinician more clearly defined by briefly

elaborating on the historical and methodological reasons for the development of the psychometric approach. When psychological tests were originally developed, group measurements of intelligence met with early and noteworthy success, especially in military and industrial settings where individual interviewing and case histories were too expensive and time consuming. An advantage of the data-oriented intelligence tests was that they appeared to be objective, which would reduce possible interviewer bias. More important, they were quite successful in producing a relatively high number of true positives when used for classification purposes. Their predictions were generally accurate and usable. However, these facts created the early expectation that all assessments could be performed using the same method and would provide a similar level of accuracy and usefulness. Later assessment strategies often tried to imitate the methods of earlier intelligence tests for variables such as personality and psychiatric diagnosis.

A further development consistent with the psychometric approach was the strategy of using a “test battery.” It was reasoned that if a single test could produce accurate descriptions of an ability or trait, administering a series of tests could create a total picture of the person. The goal, then, was to develop a global yet definitive description for the person using purely objective methods. This goal encouraged the idea that the tool (psychological test) was the best process for achieving the goal, rather than being merely one technique in the overall assessment procedure. Behind this approach were the concepts of *individual differences* and *trait psychology*. These concepts assume that one of the best ways to describe the differences among individuals is to measure their strengths and weaknesses with respect to various traits. Thus, the clearest approach to the study of personality involved developing a relevant taxonomy of traits and then creating tests to measure those traits. Again, there was an emphasis on the tools as primary, with a deemphasis on the input of the clinician. These trends created a bias toward administration and clerical skills. In this context, the psychometrist requires little, if any, clinical expertise other than administering, scoring, and interpreting tests. According to such a view, the most preferred tests would be highly standardized and ideally machine-scored so that the normed scores, rather than the psychometrist, provide the interpretation.

The objective psychometric approach is most appropriately applicable to ability tests such as those measuring intelligence or mechanical skills. Its usefulness decreases, however, when users attempt to assess personality traits such as dependence, authoritarianism, or anxiety. Personality variables are far more complex and, therefore, need to be validated in the context of history, behavioral observations, and interpersonal relationships. For example, a moderately elevated score on a scale measuring high energy level takes on an entirely different meaning for a high-functioning physician than for an individual with a history of mood disorders and associated work and interpersonal difficulties. When the purely objective psychometric approach is used for the evaluation of problems in living (coping more effectively, resolving interpersonal relationships, etc.), its usefulness is questionable. Scores need to be connected to each other and to the context in which they emerge.

Psychological assessment is most useful in the understanding and evaluation of personality and in elucidating the likely underlying causes of problems in living. These issues involve a particular problem situation having to do with a specific individual. The central role of the clinician performing psychological assessment is that of an

expert in human behavior who must deal with complex processes and understand test scores in the context of a person's life. The clinician must have knowledge concerning problem areas and, on the basis of this knowledge, form a general idea regarding behaviors to observe and areas in which to collect relevant data. Doing this involves an awareness and appreciation of multiple causation, interactional influences, and multiple relationships. As Woody (1980) stated, "Clinical assessment is individually oriented, but it always considers social existence; the objective is usually to help the person solve problems."

In addition to an awareness of the role suggested by psychological assessment, clinicians should be familiar with core knowledge related to measurement and clinical practice. This includes descriptive statistics, reliability (and measurement error), validity (and the meaning of test scores), normative interpretation, selection of appropriate tests, administration procedures, variables related to diversity (ethnicity, race, age, gender, culture, etc.), testing individuals with disabilities, and an appropriate amount of supervised experience (Turner, DeMers, Fox, & Reed, 2001). Persons performing psychological assessment should also have basic knowledge related to the demands, types of referral questions, and expectations of various contexts—particularly employment, education, vocational/career, health care (psychological, psychiatric, medical), and forensic. Furthermore, clinicians should know the main interpretive hypotheses in psychological testing and be able to identify, sift through, and evaluate a series of hypotheses to determine which are most relevant and accurate. Rather than merely knowing the labels and definitions for various types of anxiety or thought disorders, for example, clinicians should also have in-depth operational criteria for them. As another example, the concept of intelligence, as represented by the IQ score, can sometimes appear misleadingly straightforward. Intelligence test scores can be complex, though, involving a variety of cognitive abilities, the influence of cultural factors, varying performance under different conditions, and issues related to the nature of intelligence. Unless clinicians are familiar with these areas, they are not adequately prepared to handle IQ data.

The above knowledge should be integrated with relevant general coursework, including abnormal psychology, the psychology of adjustment, theories of personality, clinical neuropsychology, psychotherapy, and basic case management. A problem in many training programs is that, although students frequently have knowledge of abnormal psychology, personality theory, and test construction, they usually have insufficient training to integrate their knowledge into the interpretation of test results. Their training focuses on developing competency in administration and scoring rather than on knowledge relating to what they are testing.

The approach in this book is consistent with that of psychological assessment: Clinicians should be not only knowledgeable about traditional content areas in psychology and the various contexts of assessment but also able to integrate the test data into a relevant description of the person. This description, although focusing on the individual, should take into account the complexity of his or her social environment, personal history, and behavioral observations. Yet the goal is not merely to describe the person but rather to develop relevant answers to specific questions and present clear, specific, and reasonable recommendations that aid in problem solving and facilitate decision making.

PATTERNS OF TEST USAGE IN CLINICAL ASSESSMENT

Psychological assessment is crucial to the definition, training, and practice of professional psychology. Although the data are old, Watkins et al. (1995) found that fully 91% of all practicing psychologists engage in assessment, and 64% of all nonacademic advertisements listed assessment as an important prerequisite (Kinder, 1994). Assessment skills are also strong prerequisites for internships and postdoctoral training. The theory and instruments of assessment can be considered the very foundation of clinical investigation, applied research, and program evaluation. In many ways, psychological assessment is professional psychology's unique contribution to the wider arena of clinical practice. The early professional psychologists even defined themselves largely in the context of their role as psychological testers. Practicing psychologists spend 10% to 25% of their time conducting psychological assessment (Camara et al., 2000; Watkins, 1991; Watkins et al., 1995).

Although assessment has always been a core, defining feature of professional psychology, the patterns of use and relative importance of assessment have changed with time. During the 1940s and 1950s, psychological testing was frequently the single most important activity of professional psychologists. In contrast, the past 60 years have seen psychologists become involved in a far wider diversity of activities. Lubin and his colleagues (Lubin, Larsen, & Matarazzo, 1984; Lubin, Larsen, Matarazzo, & Seever, 1985, 1986) found that the average time spent performing assessment across five treatment settings was 44% in 1959, 29% in 1969, and only 22% in 1982. The average time spent in 1982 performing assessments in the five different settings ranged from 14% in counseling centers to 31% in psychiatric hospitals (Lubin et al., 1984, 1985, 1986). Camara et al. (2000) found that the vast majority of professional psychologists (81%) spend 0 to 4 hours a week conducting formal assessment, 15% spend 5 to 20 hours a week, and 4% spend more than 20 hours. It is expected that over the last 20 years, the time spent doing assessment has likely decreased even further. The gradual decrease in the total time spent in assessment is due in part to the widening role of psychologists. Whereas in the 1940s and 1950s a practicing psychologist was almost synonymous with a tester, professional psychologists currently are increasingly involved in administration, consultation, organizational development, and many areas of direct treatment (Bamgbose, Smith, Jesse, & Groth-Marnat, 1980; Groth-Marnat, 1988; Groth-Marnat & Edkins, 1996). Decline in testing has also been attributed to disillusionment with the testing process based on criticisms about the reliability and validity of many assessment devices (Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Wood, Lilienfeld, Garb, & Nezworski, 2000; Ziskin & Faust, 2008) and reductions in reimbursement (Cashel, 2002). In addition, psychological assessment has come to include a wide variety of activities beyond merely the administration and interpretation of traditional tests. These include conducting structured and unstructured interviews, behavioral observations in natural settings, observations of interpersonal interactions, neuropsychological assessment, behavioral assessment, and using assessment findings as part of the overall therapeutic process (Finn, 2007; Garb, 2007).

The relative popularity of different traditional psychological tests has been surveyed since 1935 in many settings, such as academic institutions, psychiatric hospitals, counseling centers, Veterans Administration centers, institutions for those with developmental disabilities, private practice, and various memberships and

professional organizations. Surveys (somewhat dated) of test usage have usually found that the 10 most frequently used tests are the Wechsler intelligence scales, Minnesota Multiphasic Personality Inventory, Rorschach, Bender Visual Motor Gestalt Test, Thematic Apperception Test, projective drawings (Human Figure Drawing, House-Tree-Person), Wechsler Memory Scale, Beck Depression Inventory, Millon Clinical Multiaxial Inventories, and California Psychological Inventory (Camara et al., 2000; Kamphaus, Petoskey, & Rowe, 2000; Lubin et al., 1985; C. Piotrowski & Zalewski, 1993; Watkins, 1991; Watkins et al., 1995). The pattern for the 10 most popular tests has remained quite stable since 1969, except that the ranking of Human Figure Drawings dropped (Camara et al., 2000). It is expected that some newer measures, especially the Personality Assessment Inventory, would be ranked quite highly in use. However, no recent surveys of test usage have been published. The pattern of test usage varies somewhat across different studies and varies considerably from setting to setting. Schools and centers for those with intellectual disabilities emphasize tests of intellectual abilities, such as the WISC-V and behavior rating scales; counseling centers are more likely to use vocational interest inventories; and psychiatric settings emphasize tests assessing level of pathology, such as the MMPI or MCMI.

One clear change in testing practices has been a relative decrease in the use and status of projective techniques (Groth-Marnat, 2000b; C. Piotrowski, 1999). Criticisms have been wide ranging but have centered on overly complex scoring systems, questionable norms, subjectivity of scoring, poor predictive utility, and inadequate or even nonexistent validity (Garb, 2005a; Garb et al., 2001; D. N. Miller, 2007; Pruitt, Smith, Thelen, & Lubin, 1985; D. Smith & Dumont, 1995). Further criticisms include the extensive time required to effectively learn the techniques, heavy reliance of projective techniques on psychoanalytic theory, and the greater time and cost efficiency of alternative objective tests. These criticisms have usually occurred from within the academic community, where the techniques are used less and less for research purposes (C. Piotrowski, 1999; C. Piotrowski & Zalewski, 1993; Watkins, 1991). As a result of these criticisms, there has been a slight but still noteworthy reduction in the use of the standard projective tests in professional practice (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Camara et al., 2000; Kamphaus et al., 2000; C. Piotrowski, 1999). Although there has been a reduction, the Rorschach and Thematic Apperception Test (TAT) continue to have a strong foothold in clinical practice. This can be attributed to lack of time available for practitioners to learn new techniques, expectations that students in internships know how to use them, unavailability of other practical alternatives, and the fact that practitioners usually give more weight to clinical experience than to empirical evidence. This suggests distance between the quantitative, theoretical world of the academic and the practical, problem-oriented world of the practitioner. In fact, assessment practices in many professional settings seem to have little relationship to the number of research studies done on assessment tools, attitudes by academic faculty, or the psychometric quality of the test (Garb, Wood, Lilienfeld, & Nezworski, 2002). In contrast to the continued use of projective instruments in adult clinical settings, psychologists in child settings are likely to rely more on behavior rating scales (e.g., Child Behavior Checklist) than projective tests (Cashel, 2002; Kamphaus et al., 2000; D. N. Miller, 2007).

The earliest form of assessment was through clinical interview. Clinicians like Freud, Jung, and Adler used unstructured interaction to obtain information regarding history, diagnosis, and underlying structure of personality. Later clinicians organized

interviews using outlines of the areas that should be discussed. During the 1960s and 1970s, much criticism was directed toward the interview, leading many psychologists to perceive interviews as unreliable and lacking empirical validation. Tests, in many ways, were designed to counter the subjectivity and bias of interview techniques. During the 1980s and 1990s, a wide variety of structured interview techniques gained popularity and have often been found to be reliable and valid indicators of a client's level of functioning. Structured interviews such as the Diagnostic Interview Schedule (DIS; Robins, Helzer, Cottler, & Goldring, 1989), Structured Clinical Interview for the DSM (SCID; Spitzer, Williams, & Gibbon, 1987), and Renard Diagnostic Interview (Helzer, Robins, Croughan, & Welner, 1981) are often given preference over psychological tests. These interviews, however, are very different from the traditional unstructured approaches. They have the advantage of being psychometrically sound even though they might lack important elements of rapport, idiographic richness, and flexibility that characterize less structured interactions (Garb, 2007; R. Rogers, 2001).

A further trend has been the development of neuropsychological assessment (see Groth-Marnat, 2000a; Lezak, Howieson, Bigler, & Tranel, 2012). The discipline is a synthesis between behavioral neurology and psychometrics and was created from a need to answer questions such as the nature of a person's organic deficits, severity of deficits, localization, and differentiating between functional versus organic impairment. The pathognomonic sign approach and the psychometric approaches are two clear traditions that have developed in the discipline. Clinicians relying primarily on a pathognomonic sign approach are more likely to interpret specific behaviors such as perseverations or weaknesses on one side of the body, which are highly indicative of the presence and nature of organic impairments. These clinicians tend to rely on the tradition of assessment associated with Luria (Bauer, 2000; Luria, 1973) and base their interview design and tests on a flexible method of testing possible hypotheses for different types of impairment. In contrast, the more quantitative tradition represented by Reitan and his colleagues (Reitan & Wolfson, 1993; Russell, 2000) is more likely to rely on critical cutoff scores, which distinguish between normal persons and those with brain damage. Reitan and Wolfson (1985, 1993) have recommended using an impairment index, which is the proportion of brain-sensitive tests that fall into the brain-damaged range. In actual practice, most clinical neuropsychologists are more likely to combine the psychometric and pathognomonic sign approaches (Rabin, Barr, & Burton, 2005). The two major neuropsychological test batteries are the Luria-Nebraska Neuropsychological Battery (Golden, Purisch, & Hammeke, 1985) and the Halstead Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1993). A typical neuropsychological battery might include tests specifically designed to assess organic impairment along with tests such as the MMPI, Wechsler intelligence scales, and the Wide Range Achievement Test (WRAT-4). As a result, extensive research over the past 15 to 20 years has been directed toward developing a greater understanding of how the older and more traditional tests relate to different types and levels of cerebral dysfunction.

During the 1960s and 1970s, behavior therapy was increasingly used and accepted. Initially, behavior therapists were concerned with an idiographic approach to the functional analysis of behavior. As their techniques became more sophisticated, formalized methods of behavioral assessment began to arise. These techniques arose in part from

dissatisfaction with the methods of diagnosis of the second edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-II*; American Psychiatric Association, 1968), as well as from a need to have assessment relate more directly to treatment and its outcomes. There was also a desire to be more accountable for documenting behavior change over time. For example, if behaviors related to anxiety decreased after therapy, the therapist should be able to demonstrate that the treatment had been successful. Behavioral assessment could involve measurements of movements (behavioral checklists, behavioral analysis), physiological responses (galvanic skin response [GSR], electromyograph [EMG]) or self-reports (self-monitoring, Symptom Checklist-90-R, assertiveness scales). Whereas the early behavioral assessment techniques showed little concern with the psychometric properties of their instruments, there has been an increasing push to have them meet adequate levels of reliability and validity (First, Frances, Widiger, Pincus, & Davis, 1992; Follette & Hayes, 1992). Despite the many formalized techniques of behavioral assessment, many behavior therapists feel that an unstructured, idiographic approach is most appropriate.

Traditional means of assessment, then, have decreased because of an overall increase in other activities of psychologists and an expansion in the definition of assessment. Currently, a psychologist doing assessment might include such techniques as interviewing, administering, and interpreting traditional psychological tests (MMPI-2/MMPI-A/MMPI-2-RF, WAIS-IV, etc.), naturalistic observations, neuropsychological assessment, and behavioral assessment. In addition, professional psychologists might be required to assess areas that were not given much emphasis before the 1980s: personality disorders (borderline personality, narcissism), stress and coping (life changes, burnout, existing coping resources), hypnotic responsiveness, psychological health, adaptation to new cultures, changes associated with increasing modernization, and strengths (related to positive psychology movements). Additional areas might include family systems interactions, relation between a person and his or her environment (social climate, social supports), cognitive processes related to behavior disorders, and level of personal control and self-efficacy. All these require clinicians to be continually aware of new and more specific assessment devices and to maintain flexibility in the approaches they take.

The future of psychological assessment will probably be most influenced by the trends toward computerized assessment, adaptation to managed health care, and distance health care delivery (Groth-Marnat, 2000b, 2009; Kay, 2007). Computerized assessment is likely to enhance efficiency through rapid scoring, complex decision rules, reduction in client–practitioner contact, novel presentation of stimuli (i.e., virtual reality), and generation of interpretive hypotheses (Lichtenberger, 2006). Future assessments are also likely to tailor the presentation of items based on the client's previous responses (Forbey & Ben-Porath, 2007). Unnecessary items will not be given, with one result being that a larger amount of information will be obtained through the presentation of relatively fewer items. This time efficiency is in part stimulated by the cost-savings policies of managed care, which require psychologists to demonstrate the cost-effectiveness of their services (Groth-Marnat, 1999; Groth-Marnat & Edkins, 1996). In assessment, this means linking assessment with treatment planning. Thus, psychological reports of the future are likely to need to link client dynamics directly to recommendations and treatment options. Whereas considerable evidence supports the

cost-effectiveness of using psychological tests in organizational contexts, health care needs to demonstrate that assessment can increase the speed of treatment as well as optimize treatment outcome (Blount et al., 2007; Groth-Marnat, 1999; Groth-Marnat, Roberts, & Beutler, 2001; Lambert & Hawkins, 2004; Yates & Taub, 2003).

A further challenge and area for development is the role distance health will play in assessment (Leigh & Zaylor, 2000; M. J. Murphy, Levant, Hall, & Glueckauf, 2007). Distance assessment as a means in and of itself is likely to become important. Professional psychologists may be required to change their traditional face-to-face role to one of developing and monitoring new applications as well as consulting/collaborating with clients regarding the results of assessments derived from the computer.

EVALUATING PSYCHOLOGICAL TESTS

Before using a psychological test, clinicians should investigate and understand the theoretical orientation of the test, practical considerations, the appropriateness of the standardization sample, and the adequacy of its psychometric properties (reliability and validity). Often, helpful descriptions and reviews that relate to these issues can be found in the test manuals as well as past and future editions of the *Mental Measurements Yearbook* (Carlson, Geisinger, & Jonson, 2014); *Tests in Print* (L. L. Murphy, Geisinger, Carlson, & Spies, 2011); *Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business* (Maddox, 2003); and *Measures for Clinical Practice: A Sourcebook* (Fischer & Corcoran, 2007). Reviews can also be found in assessment-related journals, such as the *Journal of Personality Assessment*, the *Journal of Psychoeducational Assessment*, and *Educational and Psychological Measurement*. Table 1.1 outlines the more important questions that should be answered. Each issue outlined in this table is discussed further. The discussion reflects a practical focus on problems that clinicians using psychological tests are likely to confront. It is not intended to provide a comprehensive coverage of test theory and construction; if a more detailed treatment is required, the reader is referred to one of the many texts on psychological testing (e.g., Aiken & Groth-Marnat, 2006; R. M. Kaplan & Saccuzzo, 2005).

Theoretical Orientation

Before clinicians can effectively evaluate whether a test is appropriate, they must understand its theoretical orientation. Clinicians should research the construct that the test is supposed to measure and then examine how the test approaches this construct. This information can usually be found in the test manual. If for any reason the information in the manual is insufficient, clinicians should seek it elsewhere. Clinicians can often obtain additional useful information regarding the construct being measured by carefully studying the individual test items. Usually the manual provides an individual analysis of the items, which can help the potential test user evaluate whether they are relevant to the trait being measured.

Practical Considerations

A number of practical issues relate more to the context and manner in which the test is used than to its construction. First, tests vary in terms of the level of education

Table 1.1 Evaluating a Psychological Test

Theoretical Orientation

1. Do you adequately understand the theoretical construct the test is supposed to be measuring?
2. Do the test items correspond to the theoretical description of the construct?

Practical Considerations

1. If reading is required by the examinee, does his or her ability match the level required by the test?
2. How appropriate is the length of the test?

Standardization

1. Is the population to be tested similar to the population the test was standardized on?
2. Was the size of the standardization sample adequate?
3. Have specialized subgroup norms been established?
4. How adequately do the instructions permit standardized administration?

Reliability

1. Are reliability estimates sufficiently high (generally around .90 for clinical decision making and around .70 for research purposes)?
2. What implications do the relative stability of the trait, the method of estimating reliability, and the test format have on reliability?

Validity

1. What criteria and procedures were used to validate the test?
 2. Will the test produce accurate measurements in the context and for the purpose for which you would like to use it?
-

(especially reading skill) that examinees must have to understand them adequately. The examinee must be able to read, comprehend, and respond appropriately to the test. Second, some tests are too long, which can lead to a loss of rapport with or extensive frustration on the part of the examinee. Administering short forms of the test may reduce these problems, provided these forms have been properly developed and are interpreted with appropriate caution. Finally, clinicians have to assess the extent to which they need training to administer and interpret the instrument. If further training is necessary, a plan must be developed for acquiring this training.

Standardization

Another central issue relates to the adequacy of norms (see Cicchetti, 1994). Each test has norms that reflect the distribution of scores by a standardization sample. The basis on which individual test scores have meaning relates directly to the similarity between the individual being tested and the sample. If a similarity exists between the group or individual being tested and the standardization sample, adequate comparisons can be made. For example, if the test was standardized on white American college students between the ages of 18 and 22, useful comparisons can be made for college students in

that racial and age bracket (if we assume that the test is otherwise sufficiently reliable and valid). The more dissimilar the person is from this standardization group (e.g., different national group, over 70 years of age), the less useful the test is for evaluation. The examiner may need to consult the literature to determine whether research that followed the publication of the test manual has developed norms for different groups. This is particularly important for tests such as the MMPI and the Rorschach, for which norms for various cross-national populations have been published.

Three major questions that relate to the adequacy of norms must be answered. The first is whether the standardization group includes representation from the population on which the examiner would like to use the test. The test manual should include sufficient information to determine the representativeness of the standardization sample. If this information is insufficient or in any way incomplete, it greatly reduces the degree of confidence with which clinicians can use the test. The ideal and current practice is to use stratified random sampling. However, because this can be an extremely costly and time-consuming procedure, many tests do not meet this standard. The second question is whether the standardization group is large enough. If the group is too small, the results may not give stable estimates because of too much random fluctuation. Finally, a test may have specialized subgroup norms as well as broad national norms. Knowledge relating to subgroup norms gives examiners greater flexibility and confidence if they are using the test with similar subgroup populations (see Dana, 2005). This is particularly important when subgroups produce sets of scores that are significantly different from the normal standardization group. These subgroups can be based on factors such as ethnicity, sex, geographic location, age, level of education, socioeconomic status, urban versus rural environment, or even diagnostic history. Knowledge of each of these subgroup norms allows for a more appropriate and meaningful interpretation of scores.

Standardization can also refer to administration procedures. A well-constructed test should have clear instructions that permit examiners to give the test in a manner similar to that of other examiners and also similar to themselves from one testing session and the next. Research has demonstrated that varying the instructions between one administration and the next can alter the types and quality of responses the examinee gives, thereby compromising the test's reliability. Standardization of administration should refer not only to consistent administration procedures but also to ensuring adequate lighting, quiet, no interruptions, and good rapport.

Reliability

The reliability of a test refers to its degree of stability, consistency, and predictability. It addresses the extent to which scores obtained by a person are or would be the same if the person is reexamined by the same test on different occasions. Underlying the concept of reliability is the possible range of error, or error of measurement, of a single score. This is an estimate of the range of possible random fluctuation that can be expected in an individual's score. Because psychological constructs cannot be measured directly (e.g., through measuring a level in blood), test scores are at best an approximation of these constructs, and thus error is always present in the system. It may arise from such factors as a misreading of the items, poor administration procedures, or the

changing mood of the client. If there is a large degree of error, the examiner cannot place a great deal of confidence in an individual's scores. The goal of a test constructor is to reduce, as much as possible, the degree of measurement error. If this error reduction is achieved, the difference between one score and another for a measured characteristic is more likely to result from some true difference than from some chance fluctuation.

Two main issues relate to the degree of error in a test. The first is the inevitable, natural variation in human performance. Typically variability is less for measurements of ability than for those of personality and state of being. Whereas ability variables (intelligence, mechanical aptitude, etc.) may show gradual changes resulting from growth and development, many personality traits and states of being are much more highly dependent on factors such as mood. This is particularly true in the case of a characteristic such as anxiety. The practical significance of this in evaluating a test is that certain factors outside the test itself can serve to reduce the reliability that the test can realistically be expected to achieve. Thus, an examiner should generally expect higher reliabilities for an intelligence test than for a test measuring a personality variable such as anxiety. It is the examiner's responsibility to know what is being measured, especially the degree of variability to be expected in the measured trait.

The second important issue relating to reliability is that psychological testing methods are necessarily imprecise. For the hard sciences, researchers can make direct measurements, such as the concentration of a chemical solution, the relative weight of one organism compared with another, or the strength of radiation. In contrast, many constructs in psychology are often measured indirectly. For example, intelligence cannot be perceived directly; it must be inferred by measuring behavior that has been defined as being intelligent. Variability relating to these inferences is likely to produce a certain degree of error resulting from the lack of precision in defining and observing inner psychological constructs. Variability in measurement also occurs simply because people have true (not because of test error) fluctuations in performance between one testing session and the next. Whereas it is impossible to control for the natural variability in human performance, adequate test construction can attempt to reduce the imprecision that is a function of the test itself. Natural human variability and test imprecision make the task of measurement extremely difficult. Although some error in testing is inevitable, the goal of test construction is to keep testing errors within reasonably accepted limits. A high measure of reliability is generally .80 or more, but the variable being measured also changes the expected strength of the statistic. Likewise, the method of determining reliability alters the relative strength of the statistic. Ideally, clinicians should hope for reliability statistics of .90 or higher in tests that are used to make decisions about individuals, whereas a reliability of .70 or more is generally adequate for research purposes.

The purpose of reliability is to estimate the degree of test variance caused by error. The four primary methods of obtaining reliability involve determining (1) the extent to which the test produces consistent results upon retesting (test-retest), (2) the relative accuracy of a test at a given time (alternate forms), (3) the internal consistency of the items (split-half and coefficient alpha), and (4) the degree of agreement between two examiners (interscorer). Another way to summarize this is that reliability can be time to time (test-retest), form to form (alternate forms), item to item (split-half/coefficient

alpha), or scorer to scorer (interscorer). Although these are the main types of reliability, there is a fifth type, the Kuder-Richardson; like the split-half and coefficient alpha, it is a measurement of the internal consistency of the test items. However, because this method is considered appropriate only for tests that are relatively pure measures of a single variable, it is not covered in this book.

Test-Retest Reliability

Test-retest reliability is determined by administering the test and then repeating it on a second occasion. The reliability coefficient is calculated by correlating the scores obtained by the same person on the two different administrations. The degree of correlation between the two scores indicates the extent to which the test scores can be generalized from one situation to the next. If the correlations are high, the results are less likely to be caused by random fluctuations in the condition of the examinee or the testing environment. Thus, when the test is being used in actual practice, the examiner can be relatively confident that differences in scores are the result of an actual change in the trait being measured rather than error.

A number of factors must be considered in assessing the appropriateness of test-retest reliability. One is the potential for practice and memory of a test taken on one occasion to affect performance on a second occasion, termed practice effect. Some tasks can simply improve between one administration and the next because of practice. This is a particular problem for speeded and memory tests, such as those found on the Coding and Arithmetic subtests of the WAIS-IV. Another factor to consider is that the interval between administrations, which can affect reliability. A test manual should specify the time interval, as well as any likely significant life changes that the examinees may have experienced, such as counseling, career changes, or psychotherapy. For example, tests of preschool intelligence often give reasonably high correlations if the second administration is within several months of the first one. However, correlations with later childhood or adult IQ are generally low because of innumerable, unavoidable intervening life changes. Additional sources of variation may be the result of random, short-term fluctuations in the examinee or of variations in the testing conditions. In general, test-retest reliability is the preferred method only if the variable being measured is relatively stable. If the variable is highly changeable (e.g., anxiety), this method is usually not adequate.

Alternate Forms

The alternate forms method avoids many of the problems encountered with test-retest reliability. The logic behind alternate forms is that, if the trait is measured several times on the same individual by using parallel forms of the test, the different measurements should produce similar results. The degree of similarity between the scores represents the reliability coefficient of the test. As in the test-retest method, the interval between administrations should always be included in the manual, as well as a description of any likely significant intervening life experiences. If the second administration is given immediately after the first, the resulting reliability is more a measure of the correlation between forms and not across occasions. Correlations determined by tests given with a wide time interval, such as two months or more, provide a measure of both the relation between forms and the degree of temporal stability.

The alternate forms method eliminates many carryover effects, such as the recall of specific items. However, there is still likely to be some carryover effect in that the examinee can learn to adapt to the overall style of the test even when the specific item content between one test and another is unfamiliar. This is most likely when the test involves some sort of problem-solving strategy in which the same principle in solving one problem can be used to solve the next one. An examinee, for example, may learn to use mnemonic aids to increase his or her performance on an alternate form of the WAIS-IV Digit Span subtest.

Perhaps the primary difficulty with alternate forms lies in determining whether the two forms are actually equivalent. For example, if one test is more difficult than its alternate form, the difference in scores may represent actual differences in performance on the two tests rather than differences resulting from the unreliability of the measure. Because the test constructor is attempting to measure the reliability of the test itself and not the differences between the tests, the difference between test scores could confound and lower the reliability coefficient. Alternate forms should be independently constructed tests that use the same specifications, including the same number of items, type of content, format, and manner of administration.

A final difficulty is encountered because of personal examinee differences between one administration and the next. If the alternate forms are administered on different days, the examinee may perform differently because of short-term fluctuations such as mood, stress level, or the relative quality of the previous night's sleep. Thus, an examinee's abilities may vary somewhat from one examination to another, thereby affecting test results. Despite these problems, alternate forms reliability has the advantage of at least reducing, if not eliminating, many carryover and practice effects of the test-retest method. A further advantage is that the alternate test forms can be useful for other purposes, such as assessing the effects of a treatment program (used as pre- and posttests) or monitoring a patient's changes over time by administering the different forms on separate occasions.

Internal Consistency: Split-Half Reliability and Coefficient Alpha

The split-half method and coefficient alpha are the best techniques for determining reliability for a trait with a high degree of fluctuation. Because the test is given only once and the items are correlated with each other, there is only one administration, and it is not possible for the effects of time to intervene as they might with the test-retest method. Thus, the split-half method and coefficient alpha give measures of the internal consistency of the test items rather than the temporal stability of different administrations of the same test. To determine split-half reliability, the test is often split on the basis of odd and even items. This method is usually adequate for most tests. Dividing the test into a first half and second half can be effective in some cases but is often inappropriate because of the cumulative effects of warming up, fatigue, and boredom, all of which can result in different levels of performance on the first half of the test compared with the second. This technique also would not work on a test on which items get progressively harder as the test goes on. In contrast, coefficient alpha correlates the items with each other to determine their consistency.

As is true with the other methods of obtaining reliability, the split-half method and coefficient alpha have limitations. When a test is split in half, there are fewer items on

each half, which results in wider variability because the individual responses cannot stabilize as easily around a mean. As a general principle, the longer a test is, the more reliable it is because the larger the number of items, the easier it is for the majority of items to compensate for minor alterations in responding to a few of the other items.

Interscorer Reliability

For some tests, scoring is based partially on the judgment of the examiner. Because judgment may vary between one scorer and the next, it may be important to assess the extent to which reliability might be affected. This is especially true for projectives and even for some ability tests where hard scorers may produce results somewhat different from easy scorers. This variance in interscorer reliability may apply for global judgments based on test scores, such as those with brain damage versus normal, or for small details of scoring, such as whether a person has given a shading versus a texture response on the Rorschach. The basic strategy for determining interscorer reliability is to obtain a series of responses from a single client and to have these responses scored by two different individuals. A variation is to have two different examiners test the same client using the same test and then to determine how close their scores or ratings of the person are. An interscorer reliability coefficient can be calculated using a percentage agreement, a correlation, or a kappa coefficient (which takes into account how much agreement would happen by chance). Any test that requires even partial subjectivity in scoring should provide information on interscorer reliability.

Selecting Forms of Reliability

The best form of reliability is dependent on both the nature of the variable being measured and the purposes for which the test is used. If the trait or ability being measured is highly stable, the test-retest method is preferable, whereas internal consistency is more appropriate for characteristics that are highly subject to fluctuations. When using a test to make predictions, often the test-retest method is preferable because it gives an estimate of the dependability of the test from one administration to the next. This is particularly true if, when determining reliability, an increased time interval existed between the two administrations. If, on the other hand, the examiner is concerned with measuring an individual's state (e.g., current, context-bound feelings of anxiety), split-half or coefficient alpha would likely be best.

Another consideration in evaluating the acceptable range of reliability is the format of the test. Longer tests usually have higher reliabilities than shorter ones. Also, the format of the responses affects reliability. For example, a true-false format is likely to have a lower reliability than multiple choice because each true-false item has a 50% possibility of the answer matching or being correct by chance. In contrast, each question in a multiple-choice format having five possible choices has only a 20% possibility of matching or being correct by chance. A final consideration is that tests with various subtests or subscales should report the reliability for the overall test as well as for each of the subtests. In general, the overall test score has a significantly higher reliability than its subtests. For example, the overall IQ on the WAIS-IV has a higher reliability than any of the more specific and shorter subtests used to calculate the IQ. In estimating the confidence with which test scores can be interpreted, the examiner should take

into account the lower reliabilities of the subtests. For example, based on reliability alone, a Full Scale IQ on the WAIS-IV can be interpreted with more confidence than the specific subscale scores.

Most test manuals include a statistical index of the amount of error that can be expected for test scores, which is referred to as the *standard error of measurement* (SEM). The logic behind the SEM is that test scores consist of both truth and error. Thus, there is always noise or error in the system, and the SEM provides a range to indicate how extensive that error is likely to be. The range depends on the test's reliability so that the higher the reliability, the narrower the range of error. The SEM is a standard deviation score so that, for example, a SEM of 3 on an intelligence test would indicate that an individual's score has a 68% chance of being within 3 IQ points from the estimated true score. This is because the SEM of 3 represents a band extending from -1 to $+1$ standard deviations around the mean. Likewise, there would be a 95% chance that the individual's score would fall in a range within 6 points from the estimated true score. From a theoretical perspective, the SEM is a statistical index of how a person's repeated scores on a specific test are expected to fall around a normal distribution. Thus, it is a statement of the relationship among a person's obtained score, his or her theoretically true score, and the test reliability. Because it is an empirical statement of the probable range of scores, the SEM has more practical usefulness than knowledge of the test reliability. This band of error is also referred to as a *confidence interval*.

The acceptable range of reliability is difficult to identify and depends on several factors. First is the method of reliability that is used. Alternate forms are considered to give the lowest estimate of the actual reliability of a test, while split-half provides the highest estimate. Another consideration is the length of the test. As stated previously, longer tests are expected to have higher reliability coefficients than shorter tests. One way to estimate the adequacy of reliability is by comparing the reliability derived on other similar tests, whether of the same construct or a similar design. The examiner can then develop a sense of the expected levels of reliability, which provides a baseline for comparisons. For example, when evaluating a test measuring anxiety, a clinician may not know what is an acceptable level of reliability. A general estimate can be made by comparing the reliability of the test under consideration with other tests measuring the same or a similar variable. Alternatively, a clinician may look at tests similar in construction (types of questions asked, length, etc.) but measuring a different construct for comparison. The most important thing to keep in mind is that lower levels of reliability usually suggest that less confidence can be placed in the interpretations and predictions based on the test data. However, practitioners are less likely to be concerned with low statistical reliability if they have some basis (e.g., theoretical) for believing the test is a valid measure of the client's state at the time of testing. The main consideration is that a test score should not mean one thing at one time and something different at another.

Validity

The most crucial issue in test construction is validity. Whereas reliability addresses issues of consistency, validity assesses whether a test truly measures the trait it is supposed to measure. A test that is valid for clinical assessment should measure what it is intended to measure and should also produce information useful to clinicians.

A psychological test cannot be said to be valid in any abstract or absolute sense, but more practically, it must be valid in a particular context and for a specific group of people (Messick, 1995). Although a test can be reliable without being valid, the opposite is not true; a necessary prerequisite for validity is that the test must have achieved an adequate level of reliability. That is, a test cannot truly measure what it is supposed to measure if it cannot even measure the same thing each time it is administered. Thus, a valid test is one that accurately measures the variable it is intended to measure. For example, a test comprising questions about a person's musical preference might erroneously state that it is a test of creativity. The test might be reliable in the sense that if it is given to the same person on different occasions, it produces similar results each time. However, it would not be valid in that an investigation might indicate it does not correlate highly with other more valid measurements of creativity.

Establishing the validity of a test can be extremely difficult, primarily because psychological variables are usually abstract and intangible concepts, such as intelligence, anxiety, and personality. These concepts have no tangible reality, so their existence must be inferred through indirect means. In addition, conceptualization and research on constructs undergo change over time requiring that test validation go through continual refinement (G. Smith & McCarthy, 1995). In constructing a test, a test designer must follow two necessary, initial steps. First, the construct must be theoretically evaluated and described; second, specific operations (test questions) must be developed to measure it. Even when the designer has followed these steps closely and conscientiously, it is sometimes difficult to determine what the test really measures. For example, IQ tests are good predictors of academic success, but many researchers question whether they adequately measure the concept of intelligence as it is theoretically described. Another hypothetical test that, based on its item content, might seem to measure what is described as musical aptitude may in reality be highly correlated with verbal abilities. Thus, it may be more a measure of verbal abilities than of musical aptitude.

Any estimate of validity is concerned with relationships between the test and some external independently observed event. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 1999; G. Morgan, Gliner, & Harmon, 2001) list the three main methods of establishing validity as content-related, criterion-related, and construct-related.

Content Validity

During the initial construction phase of any test, the developers must first be concerned with its content validity. *Content validity* refers to the representativeness and relevance of the assessment instrument to the construct being measured. During the initial item development, the constructors must carefully consider the skills, knowledge, or content area of the variable they would like to measure. The items are then generated based on this conceptualization of the variable. At some point, it might be decided that the item content overrepresents, underrepresents, or excludes specific areas, and alterations in the items might be made accordingly. If experts on subject matter are used to determine the items, the number of these experts and their qualifications should be included in the test manual. The instructions they received and the extent of agreement between judges

should also be provided. A good test covers not only the subject matter being measured but also additional variables. For example, factual knowledge may be one criterion, but the application of that knowledge and the ability to analyze data are also important. Thus, a test with high content validity must cover all major aspects of the content area and must do so in the correct proportion.

A concept somewhat related to content validity is face validity. These terms are not synonymous, however, because content validity pertains to judgments made by experts, whereas face validity concerns judgments made by the test users. *Face validity* refers to the degree to which a test seems like it is measuring what it purports to measure. For example, a test of arithmetic with a significant collection of arithmetic math problems to solve has high face validity. One issue in face validity is client rapport. A group of potential mechanics who are being tested for basic skills in mathematics may be better served by word problems that relate to machines rather than to business transactions. However, some tests may deliberately have low face validity, in order to decrease opportunities for examinees to skew results purposely. For example, a test like the Rorschach has low face validity for measuring a construct like psychotic thinking—examinees may not realize the test is measuring this construct—specifically to make it more difficult to fake the results in a specific direction. Despite the potential importance of face validity in regard to test-taking attitudes, disappointingly few formal studies on face validity are performed and/or reported in test manuals.

In the past, content validity has been conceptualized and operationalized as being based on the subjective judgment of the test developers. As a result, it has been regarded as the least preferred form of test validation, albeit necessary in the initial stages of test development. In addition, its usefulness has been focused primarily on achievement tests (how well has this student learned the content of the course?) and personnel selection (does this applicant know the information relevant to the potential job?). More recently, content validity has been used more extensively in personality and clinical assessment (Ben-Porath & Tellegen, 2008/2011; Butcher, Graham, Williams, & Ben-Porath, 1990; Harkness, McNulty, Ben-Porath, & Graham, 2002; Millon, Grossman, & Millon, 2015). More recent use of content validity has paralleled more rigorous and empirically based approaches to establishing validity from multiple perspectives.

Criterion Validity

A second major approach to determining validity is criterion validity, which has also been called *concurrent*, *empirical*, or *predictive validity*. Criterion validity is determined by comparing test scores with some sort of performance on an outside measure. The outside measure should have a theoretical relation to the variable that the test is supposed to measure. For example, an intelligence test might be correlated with grade point average; an aptitude test, with independent job ratings; or a test of anxiety, with other tests measuring similar constructs. The relation between the two measurements is usually expressed as a correlation coefficient.

Criterion-related validity is most frequently divided into either concurrent or predictive validity. *Concurrent validity* refers to measurements taken at the same, or approximately the same, time as the test. For example, an intelligence test might be administered at the same time as assessments of a group's level of academic

achievement. *Predictive validity* refers to outside measurements that were taken some time after the test scores were derived. Thus, predictive validity might be evaluated by correlating the intelligence test scores with measures of academic achievement a year after the initial testing. Concurrent validation is often used as a substitute for predictive validation because it is simpler, less expensive, and less time consuming and because participant attrition is not an issue. However, the main consideration in deciding whether concurrent or predictive validation is preferable depends on the test's purpose. Predictive validity is most appropriate for tests used for selection and classification of personnel. This may include hiring job applicants, placing military personnel in specific occupational training programs, screening out individuals who are likely to develop emotional disorders, or identifying which category of psychiatric populations would be most likely to benefit from specific treatment approaches. These situations all require that the measurement device provide a prediction of some future outcome. In contrast, concurrent validation is preferable if an assessment of the client's current state is required rather than a prediction of what might occur to the client at some future time. The distinction can be summarized by asking "Is Mr. Jones maladjusted?" (concurrent validity) rather than "Is Mr. Jones likely to become maladjusted at some future time?" (predictive validity).

An important consideration is the degree to which a specific test can be applied to a unique work-related environment (see Hogan, Hogan, & Roberts, 1996). This consideration relates more to the social value and consequences of the assessment than the formal validity as reported in the test manual (Messick, 1995). In other words, can the test under consideration provide accurate assessments and predictions for the environment in which the examinee is working? To answer this question adequately, the examiner must refer to the manual and assess the similarity between the criteria used to establish the test's validity and the situation to which he or she would like to apply the test. For example, can an aptitude test that has adequate criterion-related validity in the prediction of high school grade point average also be used to predict academic achievement for a population of college students? If the examiner has questions regarding the relative applicability of the test, he or she may need to undertake a series of specific tasks. The first is to identify the required skills for adequate performance in the situation involved. For example, the criteria for a successful teacher may include such attributes as verbal fluency, flexibility, and good public speaking skills. The examiner then must determine the degree to which each skill contributes to the quality of a teacher's performance. Next, the examiner has to assess the extent to which the test under consideration measures each of these skills. The final step is for the examiner to evaluate the extent to which the attribute that the test measures is relevant to the skills he or she needs to predict. Based on these evaluations, the examiner can estimate the confidence that he or she places in the predictions developed from the test. This approach is sometimes referred to as *synthetic validity* because examiners must integrate or synthesize the criteria reported in the test manual with the variables they encounter in their clinical or organizational settings.

The strength of criterion validity depends in part on the type of variable being measured. Usually, intellectual or aptitude tests give relatively higher validity coefficients than personality tests because there are generally a greater number of variables influencing personality than intelligence. As the number of variables that influences the trait

being measured increases, it becomes progressively more difficult to account for them. When a large number of variables are not accounted for, the trait can be affected in unpredictable ways. This situation can create a much wider degree of fluctuation in the test scores, thereby lowering the validity coefficient. Thus, when evaluating a personality test, the examiner should not expect as high a validity coefficient as for intellectual or aptitude tests. A helpful guide is to look at the validities found in similar tests and compare them with the test being considered. For example, if an examiner wants to estimate the range of validity to be expected for the extraversion scale on the Myers Briggs Type Indicator (MBTI), he or she might compare it with the validities for similar scales found in the NEO-PI-3 and Eysenck Personality Questionnaire. The relative level of validity, then, depends both on the quality of the construction of the test and on the variable being studied.

An important consideration is the extent to which it is realistically expected that the trait being measured should predict the trait to which it is being compared. For example, the typical correlation between intelligence tests and academic performance is about .50 (Neisser et al., 1996). Because no one would say that grade point average is entirely the result of intelligence, the relative extent to which intelligence determines grade point average has to be estimated. It can be calculated by squaring the correlation coefficient and changing it into a percentage. Thus, if the correlation of .50 is squared, it comes out to 25%, indicating that 25% of academic achievement can be accounted for by IQ as measured by the intelligence test. The remaining 75% may include factors such as motivation, quality of instruction, and past educational experience. The problem facing the examiner is to determine whether 25% of the variance is sufficiently useful for the intended purposes of the test. This determination ultimately depends on the personal judgment of the examiner.

The main problem confronting criterion validity is finding an agreed-upon, definable, acceptable, and feasible outside criterion. Whereas for an intelligence test, grade point average might be an acceptable criterion, it is far more difficult to identify adequate criteria for most personality tests. Even with so-called intelligence tests, many researchers argue that it is more appropriate to consider them tests of scholastic aptitude rather than of intelligence. Yet another difficulty with criterion validity is the possibility that the criterion measure will be inadvertently biased. Referred to as *criterion contamination*, this occurs when knowledge of the test results influences an individual's later performance. For example, a supervisor in an organization who receives such information about subordinates may act differently toward a worker placed in a certain category after being tested. This situation may set up negative or positive expectations for the worker, which could influence his or her level of performance. The result is likely to artificially increase the level of the validity coefficients. To work around these difficulties, especially in regard to personality tests, a third major method must be used to determine validity.

Construct Validity

The method of construct validity was developed in part to correct the inadequacies and difficulties encountered with content and criterion approaches. Early forms of content validity relied too heavily on subjective judgment, while criterion validity was too

restrictive in working with the domains or structure of the constructs being measured. Criterion validity had the further difficulty in that there was often a lack of agreement in deciding on adequate outside criteria. The basic approach of construct validity is to build a strong case that the test measures a theoretical construct or trait. This assessment involves three general steps. Initially, the test constructor must make a careful analysis of the trait. Then the test designer must consider the ways in which the trait should relate to other variables. Finally, the test designer needs to test whether these hypothesized relationships actually exist (Foster & Cone, 1995). For example, a test measuring dominance should have a high positive correlation with the individual accepting leadership roles, a high negative correlation with measures of submissiveness, and a very low correlation to measure of some unrelated trait, like openness. Likewise, a test measuring anxiety should have a high positive correlation with individuals who are measured during an anxiety-provoking situation, such as an experiment involving some sort of physical pain. As these hypothesized relationships are verified by research studies, the case for the measure's construct validity gets stronger and the degree of confidence that can be placed in the test increases.

There is no single, best approach for determining construct validity; rather, a variety of different possibilities exists. For example, if some abilities are expected to increase with age, correlations can be made between a population's test scores and age. This method may be appropriate for variables such as general fund of knowledge or motor coordination, but it would not be applicable for most emotional measurements. Even in the measurement of fund of knowledge or motor coordination, this approach may not be appropriate beyond the age of maturity. Another method for determining construct validity is to measure the effects of experimental or treatment interventions. Thus, a posttest measurement may be taken following a period of instruction to see if the intervention affected the test scores in relation to a previous pretest measure. For example, after an examinee completes a course in arithmetic, it would be predicted that scores on a test of arithmetical ability would increase. Often correlations can be made with other tests that supposedly measure a similar variable. However, a new test that correlates too highly with existing tests may represent needless duplication, unless it incorporates some additional advantage, such as a shortened format, ease of administration, or superior predictive validity. Related to this line of validation is presenting an argument that the test method is not majorly responsible for test scores. That is, a true/false test developed to measure anxiety should have a low correlation with a true/false test used to measure food preferences. If these scores are highly related (despite being theoretically unrelated), it may be that the scores on these tests are heavily influenced by the fact that they are true/false tests rather than by the content they are supposed to be measuring.

Factor analysis is of particular relevance to construct validation because it can be used to identify and assess the relative strength of different psychological traits. Factor analysis can also be used in the design of a test to identify the primary factor or factors measured by a series of different tests. Thus, it can be used to simplify one or more tests by reducing the number of categories to a few common factors or traits. The factorial validity of a test is the relative weight or loading that a factor has on the test. For example, if a factor analysis of a measure of anxiety determined that the test was composed of three clear factors that seemed to be measuring cognitive aspects of anxiety,

affective aspects of anxiety, and physiological aspects of anxiety, the test could be considered to have factorial validity. This would be especially true if the three factors seemed to be accounting for a clear and large portion of what the test was measuring.

Another method used as a component to build construct validity is to estimate the degree of internal consistency by correlating specific subtests with the test's total score. For example, if a subtest on an intelligence test does not correlate adequately with the overall or Full Scale IQ, it should be either eliminated or altered in a way that increases the correlation. A final method for obtaining construct validity is for a test to converge or correlate highly with variables that are theoretically similar to it. The test should not only show this convergent validity but also have discriminant validity, in which it would demonstrate low correlations with variables that are dissimilar to it. Thus, scores on reading comprehension should show high positive correlations with performance in a literature class and low correlations with performance in a class involving mathematical computation.

Related to discriminant and convergent validity is the degree of sensitivity and specificity an assessment device demonstrates in identifying different categories. *Sensitivity* refers to the percentage of true positives that the instrument has identified, whereas *specificity* is the relative percentage of true negatives. A structured clinical interview might be quite sensitive in that it would accurately identify 90% of people with schizophrenia in an admitting ward of a hospital. However, it may not be sufficiently specific in that 30% of individuals without schizophrenia would be incorrectly classified as having schizophrenia (a true negative rate of 70%). The difficulty in determining sensitivity and specificity lies in developing agreed-upon, objectively accurate outside criteria for categories such as psychiatric diagnosis, intelligence, or personality traits.

As indicated by the variety of approaches discussed, no single, quick, efficient method exists for determining construct validity. Establishing construct validity is the building of a strong case, an amassing of evidence. The process is similar to testing a series of hypotheses for which the results of the studies determine the meanings that can be attached to later test scores (Foster & Cone, 1995; Messick, 1995). Almost any data can be used, including material from the content and criterion approaches. The greater the amount of supporting data, the greater is the level of confidence with which the test can be used. As a result, construct validity represents the strongest and most sophisticated approach to test validation. In many ways, all types of validity can be considered subcategories of construct validity. Construct validation involves theoretical knowledge of the trait or ability being measured, knowledge of other related variables, hypothesis testing, and statements regarding the relationship of the test variable to a network of other variables that have been investigated (G. T. Smith, 2005). Thus, construct validation is a never-ending process in which new relationships always can be verified and investigated.

VALIDITY IN CLINICAL PRACTICE

Although a test may have been found to have a high level of validity during its construction, it does not necessarily follow that the test is also valid in a specific situation with a particular client. A test can never be valid in any absolute sense because, in practice,

numerous variables might affect the test results. A serious issue, then, is the degree of validity generalization that is made. In part, this generalization depends on the similarity between the population used during various stages of test construction and the population and situation that it is being used for in practice. Validity in clinical practice also depends on the extent to which tests can work together to improve each other's accuracy. Some tests thus show incremental validity in that they improve overall accuracy in increments as increasing numbers of data sources are used. *Incremental validity*, then, refers to the ability of tests to produce information above what is already known. Another important consideration is the ability of the clinician to generate hypotheses, test these hypotheses, and blend the data derived from hypothesis testing into a coherent, integrated picture of the person (for a full discussion of this process, see Wright, 2010). Maloney and Ward (1976) refer to this latter approach to validity as *conceptual validity* because it involves creating a conceptually coherent description of the person.

Incremental Validity

For a test to be considered useful and efficient, it must be able to produce accurate results above and beyond the results that could be obtained with greater ease and less expense (Hunsley & Meyer, 2003). If equally accurate clinical descriptions could be obtained through such basic information as biographical data and knowing the referral question, there would be no need for psychological tests. Incremental validity also needs to be evaluated in relation to cost-effectiveness. A psychological test might indeed demonstrate incremental validity by increasing the relative proportions of accurate diagnoses, or hit rates, by 2%. However, practitioners need to question whether this small increase in accuracy is worth the extra time and cost involved in administering and interpreting the test. Clinicians might focus their time more productively directly toward treatment.

In the 1950s, one of the theoretical defenses for tests having low reliabilities and validities was that, when used in combination, their accuracy could be improved. In other words, results from a series of different tests could provide checks and balances to correct for inaccurate interpretations. A typical strategy used to empirically test for this was to first obtain biographical data, make interpretations and decisions based on these data, and then test their accuracy based on some outside criterion. Next, a test such as the MMPI could be given; then the interpretations and decisions based on it could likewise be assessed for accuracy. Finally, clinicians could be given both sets of data to assess any improvements in the accuracies of interpretation/decisions between either of the first two conditions and the combined information.

It would seem logical that the greater the number of tests used, the greater would be the overall validity of the assessment battery. However, research on psychological tests used in clinical practice has often demonstrated that they have poor incremental validity. An older but representative study by Kostlan (1954) on male psychiatric outpatients compared the utility of a case history, Rorschach, MMPI, and a sentence completion test. Twenty experienced clinicians interpreted different combinations of these sources of test data. Their conclusions were combined against criterion judges who used a lengthy checklist of personality descriptions. The conclusions were that, for most of the data, the clinicians were no more accurate than if they had used only age,

occupation, education, marital status, and a basic description of the referral question. The exception was that the most accurate descriptions were based on a combination of social history and the MMPI. In contrast, psychological tests have sometimes clearly demonstrated their incremental validity. S. Schwartz and Wiedel (1981) demonstrated that neurological residents gave more accurate diagnoses when an MMPI was used in combination with history, electroencephalogram (EEG), and physical exam. This was probably not so much because of a specific MMPI neurological profile but rather because the MMPI increased diagnostic accuracy by enabling the residents to rule out other possible diagnoses.

Often clinical psychologists attempt to make a series of behavioral predictions based on complex psychological tests. Although these predictions may show varying levels of accuracy, a simpler and more effective means of achieving this information might be simply to ask the clients to predict their own behaviors. In some circumstances, self-prediction has been found to be more accurate than psychological tests, whereas in others, tests have been found to be more accurate (Shrauger & Osberg, 1981). Advantages of self-assessment are that it can be time-efficient and cost-effective and can facilitate a collegial relationship between assessor and client. In contrast, difficulties are that, compared with formal testing, self-assessment may be significantly more susceptible to social desirability, attributional errors, distortions caused by poor adjustment, and the relative self-awareness of the client. These factors need to be carefully considered before the clinician decides to use self-assessment versus formal psychological tests. Although the incremental validity of using self-assessment in combination with formal testing has not been adequately researched, it would seem that this is conceptually a potentially useful strategy for future research.

Reviews of studies on incremental validity (Garb, 1998, 2003, 2005b) have provided a number of general conclusions. The addition of an MMPI to background data has consistently led to increases in validity, although the increases were quite small when the MMPI was added to extensive data. The addition of projective tests to a test battery did not generally increase incremental validity. Lanyon and Goodstein (1982) have argued that case histories are generally preferable to psychological test data. Furthermore, a single test in combination with case history data is generally as effective as a large number of tests with case history data. Some studies have found that the MMPI alone was generally preferable to a battery containing the MMPI, Rorschach, and sentence completion (Garb, 1984, 1994a, 1998, 2005b). In contrast, other studies have found that the Rorschach can add incremental validity to a test battery (G. Meyer, 1997; Weiner, 1999).

The poor demonstrated incremental validity of many of the traditional clinical tests may relate to weaknesses and unanswered questions in the research. First, few studies have looked at statistically derived predictions and interpretations based on optimal multiple cutoff scores or multiple regression equations. However, more recent research, particularly on tests like the MMPI-2 and California Personality Inventory (CPI), has emphasized this approach. For example, combined weightings on such variables as specific CPI scores, Scholastic Aptitude Test (SAT) scores, grade point average (GPA), and IQ can be combined to predict success in specific programs (e.g., Aegisdottir, White, Spengler, Maugherman, Anderson, Cook et al., 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000). Further research using this approach may yield greater incremental

validity for a wide number of assessment techniques. Second, few studies on incremental validity have investigated the ways in which different tests might show greater incremental validity in specific situations for specific populations. Instead, most research has focused on the validity of global personality descriptions, without tying these descriptions to the unique circumstances or contexts persons might be involved in. Finally, as most previous studies have focused on global personality descriptions, certain tests demonstrate greater incremental validity when predicting highly specific traits and behaviors.

Conceptual Validity

A further method for determining validity that is highly relevant to clinical practice is conceptual validity (Maloney & Ward, 1976). In contrast to the traditional methods (content validity, etc.), which are primarily concerned with evaluating the theoretical constructs in the test itself, conceptual validity focuses on individuals with their unique histories and behaviors. It is a means of evaluating and integrating test data so that the clinician's conclusions make accurate statements about the examinee. There are similarities with construct validity in that construct validity also tries to test specific hypothesized relationships between constructs. Conceptual validity is likewise concerned with testing constructs, but in this case the constructs relate to the individual rather than to the test itself.

In determining conceptual validity, the examiner generally begins with individuals for whom no constructs have been developed. The next phase is to observe, collect data, and form a large number of hypotheses. If these hypotheses are confirmed through consistent trends in the test data, behavioral observations, history, and additional data sources, the hypotheses can be considered to represent valid constructs regarding the person. The focus is on an individual in his or her specific situation, and the data are derived from a variety of sources. The conceptual validity of the constructs is based on the logic and internal consistency of the data. Unlike construct validity, which begins with previously developed constructs, conceptual validity produces constructs as its end product. Its aim is for these constructs to provide valid sources of information that can be used to help solve the unique problems that an individual may be facing.

CLINICAL JUDGMENT

Any human interaction involves mutual and continually changing perceptions. *Clinical judgment* is a special instance of perception in which the clinician attempts to use whatever sources are available to create accurate descriptions of the client. These sources may include test data, case history, medical records, personal journals, and verbal and nonverbal observations of behavior. Relevant issues and processes involved in clinical judgment include data gathering, data synthesis, the relative accuracy of clinical versus statistical/actuarial descriptions, and judgment in determining what to include in a psychological report. This sequence also parallels the process clinicians go through when assessing a client.

Data Gathering and Synthesis

Most of the research related to the strengths and weaknesses of data gathering and synthesis has focused on the assessment interview (see Chapter 3). However, many of the issues and problems related to clinical judgment during interviewing also have implications for the gathering and synthesis of test data. One of the most essential elements in gathering data from any source is the development of an optimum level of rapport. Rapport increases the likelihood that clients will give their optimum level of performance. If rapport is not sufficiently developed, it is increasingly likely that the data obtained from the person will be inaccurate.

Another important issue is that the interview itself is typically guided by the client's responses and the clinician's reaction to these responses. A client's responses might be nonrepresentative because of factors such as a transient condition (stressful day, poor night's sleep, etc.) or conscious/unconscious faking. The client's responses also need to be interpreted by the clinician. These interpretations can be influenced by a combination of personality theory, research data, and the clinician's professional and personal experience. The clinician typically develops hypotheses based on a client's responses and combines his or her observations with his or her theoretical understanding of the issue. These hypotheses can be further investigated and tested by interview questions and test data, which can result in confirmation, alteration, or elimination of the hypotheses. Thus, bias can potentially enter into this process from a number of different directions, including the types of questions asked, initial impressions, level of rapport, or theoretical perspective.

The clinician typically collects much of the initial data regarding a client through unstructured or semistructured interviews. Unstructured approaches in gathering and interpreting data provide flexibility, focus on the uniqueness of the person, and are ideographically rich. In contrast, an important disadvantage of unstructured approaches is that a clinician, like most other persons, can be influenced by a number of personal and cultural biases. For example, clinicians might develop incorrect hypotheses based on first impressions (primacy effect). They might end up seeking erroneous confirmation of incorrect hypotheses by soliciting expected responses rather than objectively probing for possible disconfirmation. Thus, clinicians might be unduly influenced by their preferred theory of personality, halo effects, expectancy bias, and cultural stereotypes. These areas of potential sources of error have led to numerous questions regarding the dependability of clinical judgment.

Accuracy of Clinical Judgments

After collecting and organizing their data, clinicians then need to make final judgments regarding the client. Determining the relative accuracy of these judgments is crucial. In some cases, clinical judgment is clearly in error, whereas in others it can be quite accurate. Cultural bias can come into play, and clinicians should take into consideration cultural context and personal beliefs when making clinical judgments. To increase accuracy, clinicians need to know how errors might occur, how to correct these errors, and the relative advantages of specialized training.

A possible source of inaccuracy is that clinicians frequently do not take into account the base rate, or the rate at which a particular behavior, trait, or diagnosis occurs in the general population (Faust, 1991; S. Hawkins & Hastie, 1990; Wedding & Faust, 1989). For example, an intake section of a psychiatric hospital might use a test that has been shown to be 90% accurate at telling whether a person has schizophrenia. Perhaps 5% of the time the test shows a false positive and 5% of the time it shows a false negative. If a person comes in and the test reveals a positive result for schizophrenia, it is not necessarily a 90% or 95% chance that he or she actually has schizophrenia. Because schizophrenia has a low base rate (e.g., if roughly 1% of the population has it), there is actually a much greater than 10% chance that this individual does not have schizophrenia.

It is also rare for clinicians to receive feedback regarding either the accuracy of their diagnoses or other frequently used judgments, such as behavioral predictions, personality traits, or the relative success of their recommendations (Garb, 1989, 1994a, 1998, 2005b). Thus, it is possible that inaccurate strategies for arriving at conclusions will continue with little likelihood of correction.

A further source of error is that information obtained earlier in the data collection process is frequently given more importance than information received later (primacy effect). This means that different starting points in the decision-making process may result in different conclusions. This error can be further reinforced if clinicians make early judgments and then work to confirm these judgments through seeking supporting information. The resulting *confirmatory bias* is especially likely to occur in a hypothesis-testing situation in which clinicians do not adequately seek information that could disconfirm as well as confirm their hypothesis (Haverkamp, 1993). The most problematic examples occur when clinicians interpret a client's behavior and then work to persuade the client that their interpretation is correct (Loftus, 1993).

Research on person perception accuracy indicates that, even though nobody is uniformly accurate, some persons are much better at accurately perceiving others. Taft (1955) and P. E. Vernon (1964) summarized the early research on person perception accuracy by pointing out that accuracy is not associated with age (in adults); there is little difference in accuracy between males and females (although females are slightly better); and accurate perceptions of others are positively associated with intelligence, artistic/dramatic interests, social detachment, and good emotional adjustment. Authoritarian personalities tend to be poor judges. In most instances, accuracy is related to similarity in race and cultural backgrounds (P. Shapiro & Penrod, 1986). In some cases, accuracy by psychologists may be only slightly related to their amount of clinical experience (Garb, 1989, 1992, 1994a, 1998, 2005b); and, for some judgments, psychologists may be no better than certain groups of nonprofessionals, such as physical scientists and personnel workers (Garb, 1992, 1994a, 1998, 2005b). Relatively higher rates of accuracy were achieved when clinical judgments based on interviews were combined with formal assessments and when statistical interpretive rules were used. When subjective test interpretation was combined with clinical judgment, it was questionable whether any increase in accuracy was obtained (Garb, 1984, 1989).

It would be logical to assume that the more confidence clinicians feel regarding the accuracy of their judgments, the more likely it is that their judgments are accurate. In several studies, however, confidence was often not related to accuracy (E. Kelly &

Fiske, 1951; Kleinmuntz, 1990). Kelly and Fiske even found that degree of confidence was inversely related to predicting the success of trainees in a Veterans Administration training program. Several studies (Kareken & Williams, 1994; Lichtenstein & Fischhoff, 1977) concluded that persons were generally overconfident regarding judgments; and when outcome knowledge was made available, clinicians typically overestimated what they thought they knew before receiving outcome knowledge (Hawkins & Hastie, 1990). This overconfidence is usually referred to as *hindsight bias* (“I would have known it all along”) and is usually accompanied by a denial that the outcome knowledge has influenced judgment. Paradoxically, as knowledge and experience in an area increase, there is generally a decrease in confidence regarding judgments. This observation was found to be true unless the clinicians were very knowledgeable, in which case they were likely to have a moderate level of confidence (Garb, 1989). Confidence was also higher if participants were made socially accountable for their judgments (Ruscio, 2000). Thus, the more experienced clinicians and persons who were more socially accountable rated their level of confidence as higher.

Crucial to clinical judgment is whether clinicians can make judgments better than laypersons and whether amount of clinical training can increase accuracy. This is a particularly important issue if psychologists are offering their services as expert witnesses to the legal justice system. Research reviews generally support the value of clinical training, but this is dependent on the domain being assessed. For example, Garb (1992) concluded, “Clinicians are able to make reliable and valid judgments for many tasks, and their judgments are frequently more valid than judgments by laypersons” (p. 451). In particular, clinicians have been found to make more accurate judgments relating to relatively complex technical areas, such as clinical diagnosis, ratings of mental status, many domains related to interview information, short-term (and possibly long-term) predictions of violence, psychological test interpretation (WAIS, MMPI), forensic knowledge, competency evaluations, neuropsychological test results, psychotherapy data, and biographical data (see primarily Garb, 1998, but also 1984, 1989, 1992, 1994a). In contrast, trained clinicians were no better than less experienced persons (laypersons, novice trainees) in making judgments based on projective test results and in making personality descriptions based on face-to-face interaction (Garb, 2005b; Witteman & van den Bercken, 2007).

The preceding material indicates that errors in clinical judgment can and do occur. It is thus crucial, especially when appearing as an expert in court, that clinicians be familiar with the relevant literature on clinical judgment and, based on this information, take steps to improve their accuracy. Accordingly, Garb (1994a, 1998, 2005b) and Wedding and Faust (1989) made the following recommendations:

1. To avoid missing crucial information, clinicians should use comprehensive, structured, or at least semistructured approaches to interviewing. This is especially important in cases where urgent clinical decisions (danger to self or others) may need to occur.
2. Clinicians should not only consider the data that support their hypotheses, but they should also carefully consider or even list evidence that does not support their hypotheses. This method will likely reduce the possibility of hindsight and confirmatory bias.

3. Diagnoses should be based on careful attention to the specific criteria contained in the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013)* or *International Classification of Disorders (ICD-10; World Health Organization, 1992)*. In particular, this means not making errors caused by inferences biased by gender and ethnicity.
4. Because memory can be a reconstructive process subject to possible errors, clinicians should avoid relying on memory and rather refer to careful notes as much as possible.
5. In making predictions, clinicians should attend to base rates as much as possible. Such a consideration potentially provides a rough estimate of how frequently the behavior will occur in a given population or context. Any clinical predictions, then, are guided by this base rate occurrence and are likely to be improvements on the base rate.
6. Clinicians should seek feedback when possible regarding the accuracy and usefulness of their judgments. For example, psychological reports should ideally be followed up with rating forms (that can be completed by the referral sources) relating to the clarity, precision, accuracy, and usefulness of the information and recommendations contained in the reports.
7. Clinicians should learn as much as possible regarding the theoretical and empirical material relevant to the person or group they are assessing. Doing this would potentially help clinicians to develop strategies for obtaining comprehensive information, allow them to make correct estimates regarding the accuracy of their judgments, and provide them with appropriate base rate information.
8. Practitioners should be familiar with the literature on clinical judgment in order to continually update their knowledge on past and emerging trends.

Sometimes in court proceedings, psychologists are challenged regarding the difficulties associated with clinical judgment. If the preceding steps are taken, psychologists can justifiably reply that they are familiar with the literature and have taken appropriate steps to guard against inaccuracies in clinical judgment. More important, by taking these steps, the clinicians' quality of service related to clients and referral sources is also likely to be enhanced.

Clinical Versus Actuarial Prediction

Over 60 years ago, Meehl (1954) published a review of research comparing the relative accuracy of clinical judgment to statistical formulas when used on identical sets of data (life history, demographic data, test profiles). The clinical approach used clinicians' judgment, whereas the actuarial approach used empirically derived formulas, such as single/multiple cutoffs and regression equations, to come to decisions regarding a client. His review covered a large number of settings including military placement, college success, criminal recidivism, and benefit from psychotherapy. He concluded that statistical decisions consistently outperformed clinical judgments (Meehl, 1954, 1965). Some lively debate in the journals ensued, with Meehl's conclusions generally being supported (Aegisdottir et al., 2006; Garb, 1994b; Grove et al., 2000; Kleinmuntz, 1990).

The magnitude of this difference has been estimated to be a 13% greater accuracy using actuarial methods when compared with clinical judgment.

Despite the empirical support for an actuarial approach, several practical and theoretical issues need to be considered. A clinical approach to integrating data and arriving at conclusions allows a clinician to explore, probe, and deepen his or her understanding in many areas. These explorations frequently involve areas that tests or statistical formulas cannot measure. Often an interview is the only means of obtaining observations of behavior and unique aspects of history. Idiosyncratic events with a low frequency of occurrence may significantly alter a clinician's conclusions although no formulas take these events into account. It is quite common for unique, rare events to have occurred at some time in a client's life; and, during the process of assessment, they are frequently relevant and can often alter the conclusions of many, if not most, clinical assessments. Not only do unique aspects of a person change interpretations, but typically an assessment for a person needs to be focused for a specific context and specific situation that he or she is involved in. When the focus changes from institutional to individual decision making, the relevance of statistical rules becomes less practical (McGrath, 2001; Vane & Guarnaccia, 1989). Not only are individuals too multifaceted for simple actuarial formulas, but their unique situations, contexts, and the decisions facing them are even more multifaceted.

A further difficulty with a purely actuarial approach is that development of both test reliability and validity, as well as actuarial formulas, requires conceiving the world as stable and static. For such approaches to be useful, the implicit assumption is that neither people nor criteria change. In contrast, the practitioner must deal with a natural world that is imperfect, constantly changing, does not necessarily follow rules, is filled with constantly changing perceptions, and is subject to chance or at least impossible-to-predict events. Thus, even when statistical formulas are available, they may not apply. This distinction between the statistical orientation of the psychometrician and the natural environment of the practitioner underlies the discrepancy between their two worlds (Beutler, 2000). Practitioners must somehow try to combine these two modes of analysis, but they often find the task difficult. It may be true that controlled studies generally favor a statistical approach over a clinical one, but, at the same time, that truth is seldom useful to the practitioner involved in the changing and unique world of practice (Bonarius, 1984).

Bonarius (1984) presented a conceptual alternative to this dilemma. The first step is to alter mechanistic views of prediction. Instead, clinicians might avoid the term *prediction* altogether and use *anticipation*. Anticipating future possibilities implies a cognitive constructional process rather than a mechanical process. It admits that the world can never be perfect in any mechanistic sense and that there is no such thing as an average person in an average situation engaged in an average interaction. Furthermore, the creation of future events is shared by coparticipants. Clients take an active part in formulating and evaluating their goals. The success of future goals depends on the degree of effort they are willing to put into them. The coparticipants share responsibility for the future. Thus, the likelihood that future events will occur is related to both cognitive constructions of an idiosyncratic world and interaction between participants.

Ideally, clinicians need to be aware of and to use, whenever available, actuarial approaches, such as multiple cutoffs and regression equations. Doing so would be

particularly important for situations where there are clearly defined outcomes, errors are costly, and clinicians need to have maximum accountability. Such situations might include suicide, violence, sexual offending, recidivism, relapse, postparole adjustment, malingering, response to psychotherapy, academic performance, vocational success, psychiatric prognosis, or success in training programs. Despite over 50 years of research and debates, actuarial strategies are still not widely available except within forensic contexts. In addition, many of the formulas are “not ready for prime time” (Aegisdottir et al., 2006). It is hoped that at some time in the future, a set of optimal, well-validated actuarial formulas will be widely available along with user-friendly programs on how to use them (Groth-Marnat, 2000b, 2009). The results from such formulas will still need to be integrated with data and inferences obtainable only through clinical means. Although it is unlikely that actuarial prediction rules will replace clinical judgment, formal prediction rules can and should be used more extensively as a resource to improve the accuracy of clinical decision making.

Psychological Report

An accurate and effective psychological report requires that clinicians clarify their thinking and crystallize their interpretations. The report ties together all sources of information, often combining complex interprofessional and interpersonal issues. All the advantages and limitations involved with clinical judgment either directly or indirectly affect the report. The focus should be a clear communication of the clinician’s interpretations, conclusions, and recommendations. Chapter 15 provides in-depth information on the psychological report as it relates to relevant research, guidelines, format, and sample reports.

PHASES IN CLINICAL ASSESSMENT

An outline of the phases of clinical assessment can provide both a conceptual framework for approaching an evaluation and a summary of some of the points already discussed. Although the steps in assessment are isolated for conceptual convenience, in actuality, they often occur simultaneously and interact with one another. Throughout these phases, the clinician should integrate data and serve as an expert on human behavior rather than merely an interpreter of test scores. Doing so is consistent with the belief that a psychological assessment can be most useful when it addresses specific individual problems and provides guidelines for decision making regarding these problems.

Evaluating the Referral Question

Many of the practical limitations of psychological evaluations result from an inadequate clarification of the problem. Because clinicians are aware of the assets and limitations of psychological tests, and because clinicians are responsible for providing useful information, it is their duty to clarify the requests they receive. Furthermore, they cannot assume that initial requests for an evaluation are adequately stated.

Clinicians may need to uncover hidden agendas, unspoken expectations, and complex interpersonal relationships. One of the most important general requirements is that clinicians understand the vocabulary, conceptual model, dynamics, and expectations of the referral setting in which they will be working (Turner et al., 2001). Further, clinicians must evaluate whether the referral questions are appropriate for psychological assessment and whether they have a level of competence necessary to conduct an assessment to answer the specific questions.

Clinicians are rarely asked to give a general or global assessment but instead are asked to answer specific questions. To address these questions, it is sometimes helpful to contact the referral source at different stages in the assessment process. For example, it is often important in an educational evaluation to observe the student in the classroom environment. The information derived from such an observation might be relayed back to the referral source for further clarification or modification of the referral question. Likewise, an attorney may wish to somewhat alter his or her referral question based on preliminary information derived from the clinician's initial interview with the client.

Data Collection

After clarifying the referral question and obtaining knowledge related to the problem, clinicians can proceed with the actual collection of information. The information may come from a wide variety of sources, the most frequent of which are interview data, collateral information, behavioral observations, and test scores. Collateral information may include school records, previous psychological reports, medical records, police reports, or interviews with parents or teachers. It is important to realize that the tests themselves are merely a single tool, or source, for obtaining data. The case history is of equal importance because it provides a context for understanding the client's current problems and, through this understanding, renders the test scores meaningful. In many cases, a client's history is of even more significance in making predictions and in assessing the seriousness of his or her condition than his or her test scores. For example, a high score on depression on the MMPI-2 is not as helpful in assessing suicide risk as are historical factors, such as the number of previous attempts, details regarding any previous attempts, and length of time the client has been depressed. Moreover, test scores themselves are usually not sufficient to answer the referral question. For specific problem solving and decision making, clinicians must rely on multiple sources and, using these sources, check to assess the consistency of the observations they make.

Before beginning the actual testing procedure, examiners should carefully consider the problem, the adequacy of the tests they will use, and the specific applicability of that test to an individual's unique situation. This preparation may require referring both to the test manual and to additional outside sources. Clinicians should be familiar with operational definitions for problems such as anxiety disorders, psychoses, personality disorders, and organic impairment so that they can be alert to their possible expression during the assessment procedure. Clinicians should also be familiar with problems that can arise from medical conditions and substance use. Competence in merely administering and scoring tests is insufficient to conduct effective assessment. For example, the development of an IQ score does not necessarily indicate that an examiner is aware of differing cultural expressions of intelligence or of the limitations of the assessment

device. It is essential that clinicians have in-depth knowledge about the variables they are measuring; if not, their evaluations are likely to be extremely limited.

When evaluating whether a test will be useful in a specific case, a clinician should consider several factors. The relative adequacy of the test will include inquiry about certain practical considerations, the standardization sample, and reliability and validity (see Table 1.1). Specifically, a test should truly measure a construct of interest in the specific case. It is important that the examiner also consider whether a specific test or tests are appropriate to use on an individual or group. Doing this demands knowledge in such areas as the client's age, sex, ethnicity, race, culture, educational background, motivation for testing, anticipated level of resistance, social environment, and interpersonal relationships. Finally, clinicians need to assess the effectiveness or utility of the test in aiding the treatment process.

Interpreting the Data

The end product of assessment should be a set of recommendations that are clear, specific, and reasonable. In order to support these recommendations, clinicians should be able to describe the client's current level of functioning, considerations relating to etiology, and prognosis. Etiologic descriptions should avoid simplistic formulas and should instead focus on the influence exerted by several interacting factors, which may include primary, predisposing, precipitating, and reinforcing causes. Further elaborations may also attempt to assess the person from a systems perspective, in which the clinician evaluates patterns of interaction, mutual two-way influences, and the specifics of reciprocal information feedback. An additional crucial area is to use the data to develop an effective plan for intervention (see Beutler, Clarkin, & Bongar, 2000; Harwood, Beutler, & Groth-Marnat, 2011; Hersen, 2005a; Jongsma, Peterson, & Bruce, 2014; Maruish, 2004). Clinicians should also pay careful attention to research on, and the implications of, incremental validity and continually be aware of the limitations and possible inaccuracies involved in clinical judgment. If actuarial formulas are available, they should be used when possible. These considerations indicate that the description of a client should not be a mere labeling or classification but should rather provide a deeper and more accurate understanding of the person. This understanding should allow the examiner to perceive new facets of the person in terms of both his or her internal experience and his or her relationships with others.

To develop these descriptions, clinicians must make inferences from their test data. Although such data are objective and empirical, the process of developing hypotheses, obtaining support for these hypotheses, and integrating the conclusions is dependent on the theoretical knowledge and understanding, experience, and training of the clinician. This process generally follows a sequence of developing hypotheses, identifying relevant facts, making inferences, and supporting these inferences with relevant and consistent data. Wright (2010) conceptualized an eight-phase approach (Figure 1.1) for using data in a psychological assessment. It should be noted that, in actual practice, these phases are not as clearly defined as indicated in the figure, but often occur simultaneously. For example, when a clinician reads a referral question or initially observes a client, he or she is already developing hypotheses about that person and checking to assess the validity of these observations.

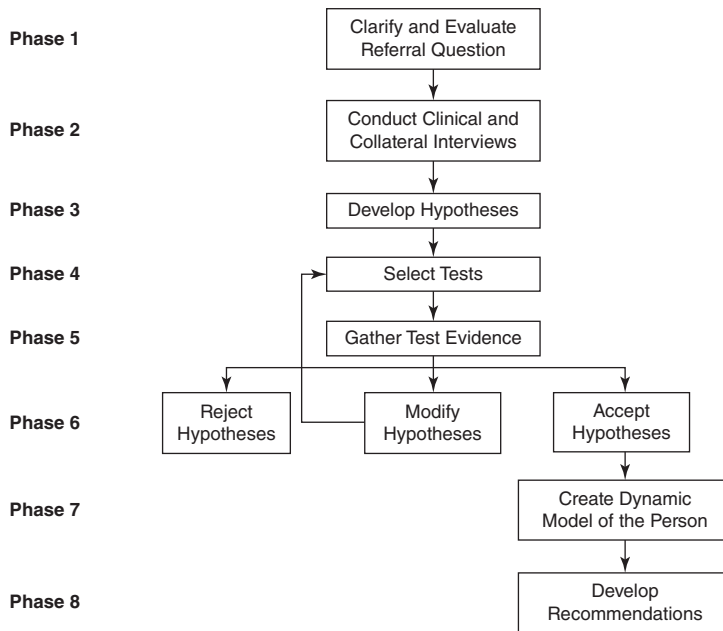


Figure 1.1 Hypothesis testing model for interpreting assessment data

Source: Adapted from Wright, 2010. Reprinted with permission from *Conducting Psychological Assessment: A Guide for Practitioners*, by A. J. Wright, Hoboken, NJ: Wiley.

Phase 1

The first phase, discussed above, is the clarification and evaluation of the referral question. As referral questions are one source of data, the clinician is already starting to develop hypotheses about what is going on for a client, what impact it has on his or her life, under what conditions the current problems developed, and even possible recommendations for how to improve the client's functioning and life in general.

Phase 2

Phase 2 focuses on collecting another source of data through clinical interviews and other background information (e.g., through collateral interviews, such as with parents or teachers, or through reviewing records or previous reports). Clinicians must understand the strengths and limitations of data collected from clinical interviews (see Chapter 3). It is from these data, though, that clearer initial hypotheses can be formed about the client's cognitive, emotional, personality, academic, neuropsychological, adaptive, and other areas of functioning.

Phase 3

Based on the information collected in Phases 1 and 2, the third phase focuses on developing hypotheses about what factors (situations, internal dynamics, etc.) may be causing and/or reinforcing whatever problems the client is having. These hypotheses require the clinician to have a firm grasp on many content areas of psychology, including personality theory, developmental psychology, abnormal psychology, developmental

neurobiology, and even areas outside of psychology like biology, sociology, and cultural anthropology.

These hypotheses must be grounded in clear and logical clinical science and theory, regardless of theoretical orientation. For example, a hypothesis about the etiology of a client's low self-esteem may revolve around negative self-talk (from a cognitive-behavioral perspective) or the internalization of a mother's criticism (from a psychodynamic perspective). Regardless of theoretical orientation, the hypothesis must make sense within a specific psychological framework.

Phase 4

The importance of deliberateness when selecting tests to use in a specific assessment battery cannot be overstated. In addition to the considerations discussed earlier (see Table 1.1), the clinician must be confident that the tests selected can rule in or out the specific hypotheses generated in Phase 3 (as well as any modified hypotheses later on). Special attention should always be paid to cultural and sociodemographic characteristics of the client in order to ensure that the tests selected are appropriate, given the development, standardization, and norming procedures of the tests being considered.

Phases 5 and 6

Phase 5 centers on administering and scoring tests in order to collect data to evaluate the hypotheses generated in Phase 3. Phase 6, one of the most difficult phases, relates to the actual evaluation of test data within the context of the hypotheses generated previously. Phases 4 through 6 are iterative and recursive. As test data are collected, hypotheses can be rejected, modified, or accepted. Rejected hypotheses are abandoned, and the clinician can confidently move on to evaluating other hypotheses. Modified hypotheses may require the selection of new tests; while some tests may help develop modified hypotheses, additional tests are often necessary to actually evaluate these new hypotheses.

While rejecting and modifying hypotheses is often relatively straightforward, accepting hypotheses can be much more difficult, especially when it comes to personality or emotional functioning. It is often the case that a test or test score can rule *out* a hypothesis but cannot rule it *in*. For example, a high score on the Working Memory Index (WMI) of the WISC-V may rule out the presence of the inattentive subtype of attention-deficit/hyperactivity disorder (ADHD). This is because a child with ADHD would find it very difficult, if not impossible, to perform extremely well on WMI tasks that require both selective and sustained attention. However, a low score on the same WMI cannot rule ADHD *in*. Because multiple factors can affect performance on the WMI, more testing would be necessary to investigate the case of whether or not ADHD was present.

Phase 7

Phase 7 is a complicated phase requiring the clinician to make sense of all of the data collected in a way that can be clearly communicated to the client and/or referral source. Rather than presenting an acontextual list of a client's strengths and weaknesses or, even worse, presenting data test by test (which requires the audience to then determine

which findings are important and connect the dots to make sense of the feedback), clinicians should create a dynamic understanding of how factors interact to explain what is happening for the client. To do this process well takes good training, supervision, and experience.

Phase 8

The final phase of the data interpretation process is linking the results to clear, specific, and reasonable recommendations that are likely to improve the client's life and functioning. Chapter 14 focuses on this process. In short, clinicians must understand treatment options from two different perspectives. First, clinically, clinicians must understand what is likely to link to and address the specific problems that emerged from the assessment, including the dynamics identified in Phase 7. Second, clinicians must understand the research behind interventions, how effective they have been shown to be, and what about them has been suggested or found to be the reasons that they are effective. Clinicians must consider both the empirical support of interventions and the likelihood of the interventions benefitting the specific client in his or her specific context and situation. Recommendations cannot be vague or broad, such as recommending "therapy" to a client. They should be both clear and specific. Additionally, they should be reasonable, given the circumstances. Although a specific treatment may be the best choice for a specific client, for a number of reasons, if that treatment is not available to the client (because of, for example, geographic location or financial limitations), then making a recommendation for that kind of treatment will not ultimately benefit the client.

RECOMMENDED READING

- Aiken, L. R., & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston, MA: Pearson Education.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Psychology*, 1, 67–89.
- Groth-Marnat, G. (2000). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology*, 56, 349–365.
- Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., . . . Eisman, E. J. (2000). *Empirical support for psychological assessment in clinical care settings. Professional Psychology*, 31, 119–130.
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist*, 45, 999–1017.
- Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., . . . Reed, G. M. (2001). Psychological testing and psychological assessment. *American Psychologist*, 56, 128–165.
- Wright, A. J. (2010). *Conducting psychological assessment: A guide for practitioners*. Hoboken, NJ: Wiley.

CONTEXT OF CLINICAL ASSESSMENT

Although general knowledge regarding tests and test construction is essential, practitioners must consider a wide range of additional issues to situate testing procedures and test scores within an appropriate context. These considerations include clarifying the referral question, understanding the referral context, following ethical guidelines, identifying and working with test bias, selecting the most appropriate instrument for the variable or problem being studied, and making appropriate use of computer-assisted interpretation.

TYPES OF REFERRAL SETTINGS

Throughout the assessment process, practitioners should try to understand the unique problems and demands encountered in different referral settings. Otherwise, examiners—despite being skilled in administering and interpreting tests—might administer a needless series of tests and, at worst, provide useless information to referral sources and patients themselves. That is, a thorough investigation of the underlying motive for a referral can sometimes lead to the discovery that evaluation through testing is not even warranted.

Errors in test interpretation frequently occur because clinicians do not respond to the referral question in its broadest context. In turn, requests for psychological testing are often worded vaguely: “I would like a psychological evaluation on Mr. Smith,” or “Could you evaluate Jimmy because he is having difficulties in school?” The request often does not state a specific question that must be answered or a decision that must be made, although many times this is the position that the referral source is in. For example, a school administrator may need testing to support a placement decision, a teacher may want to prove to parents that their child has a serious problem, or a psychiatric resident may not be comfortable with the management of a patient. An organization’s surface motive for testing may be as vague as a statement that the procedure is a matter of policy. Greater clarification is necessary before clinicians can provide useful problem-solving information. Furthermore, many of these situations have hidden agendas that may not be adequately handled through psychological testing alone. One of the most useful questions in addressing these issues is to ask what decisions need to be made regarding the patient.

It must be stressed that the responsibility for exploring and clarifying the referral question lies with the clinician, who should actively work with the referral source to place the client’s difficulty in a practicable context. Clinicians must understand the decisions that the referral source is facing, as well as the potential alternatives and

their possible implications. Clinicians also need to specify the potential utility of the psychological evaluation in determining different alternatives and their possible outcomes. They should make clear the advantages and usefulness of psychological testing but should also explain the limitations inherent in the process.

To help clarify the referral question, and to develop a relevant psychological evaluation, clinicians should become familiar with the types of environments in which they will be working. The most frequent environments are the psychiatric setting, the general medical setting, the legal context, the educational context, and the psychological clinic.

Psychiatric Setting

Levine (1981) summarized the important factors for a psychologist to be aware of in a psychiatric setting. These referrals typically come from a psychiatrist, who may be asking the referral question in the role of administrator, therapist, or physician. Each role presents unique issues for the psychiatrist, and clinicians have the primary responsibility to develop evaluations that directly address the problems at hand.

One of the main roles a psychiatrist fills is administrator on a ward. Ward administrators must frequently make decisions about problems such as suicide risk, admission/discharge, and the suitability of a wide variety of medical procedures. While retaining ultimate decision-making responsibility, psychiatrists often use information from other persons to help with decisions. This represents a change from the typical role of psychiatrists 40 years ago when they were mainly concerned with diagnosis and treatment. Currently, issues about custody, freedom of the patient, and the safety of society have taken over as the primary focus. From the perspective of psychologists performing assessments, this means that making a formal *DSM-5* (American Psychiatric Association, 2013) psychiatric diagnosis is usually not sufficient in and of itself. For example, a patient may be diagnosed bipolar, but this label does not indicate the level of dangerousness that the patient poses to him- or herself or to others. After patients have been admitted to a psychiatric setting, many practical questions have to be answered, such as the type of ward in which to place them, the activities in which they should be involved, and the method of therapy that would be most likely to benefit them.

Initially, the psychologist must determine exactly what information the ward administrator is looking for, particularly concerning any decisions that must be made about the patient. Psychologists in psychiatric settings who receive vague requests for “a psychological” sometimes develop a standard evaluation based on what they have learned about what this term implies on their specific unit. They may evaluate the patient’s defense mechanisms, diagnosis, cognitive style, and psychosocial history without addressing the specific decisions that have to be made or perhaps covering only two or three relevant issues and omitting others. To maximize the usefulness of an evaluation, examiners must be especially aware of, and sensitive to, psychiatric administrators’ legal and custodial responsibilities.

In contrast to the concerns of ward administrators, the standard referral questions from psychiatrists evaluating a patient for possible psychotherapy involve the appropriateness of the client for such therapy, the strategies that are most likely to be effective, and the likely outcome of therapy. These assessments are usually clear-cut and typically do not present many difficulties. Such evaluations can elaborate on likely problems

that may occur during the course of therapy, capacity for insight, diagnosis, coping style, level of resistance, degree of functional impairment, and problem complexity (see Chapter 14).

If a referral is made during therapy, however, a number of problem areas may exist that are not readily apparent from the referral question. The evaluator must investigate these complicating factors, along with potential decisions that may flow from the assessment information. An area of potential conflict arises when psychiatrists are attempting to fulfill roles of both administrator (caretaker) and psychotherapist and yet have not clearly defined these roles either for themselves or for their patients. The resulting ambiguity may cause the patient to feel defensive and resistant and the psychiatrist to feel that the patient is not living up to the therapist's expectations. Elaboration of a specific trait or need in the patient cannot resolve this conflict but must occur in the context of interactions between the therapist and the patient. A standard psychological evaluation investigating the internal structure of the patient will not address this issue.

A second possible problem area for clients referred in the midst of therapy can be the result of personal anxiety and discomfort on the therapist's part. Thus, issues such as therapist bias and possible unreasonable expectations may be equally or even more important than looking at a patient's characteristics. If role ambiguity, countertransference, bias, or unreasonable expectations are discovered, they must be elaborated and communicated in a sensitive manner.

When psychiatrists are acting in the role of physician, they and the psychologist may have different conceptual models for describing a patient's disorder. Whereas psychiatrists function primarily from a disease or medical model, psychologists may speak in terms of difficulties in living with people and society. In effectively communicating the results of psychological evaluations, examiners must bridge this conceptual difference. For example, a psychiatrist may ask whether a patient has a dissociative disorder, whereas a psychologist may not believe that the label dissociative disorder is useful or even a scientifically valid concept. The larger issue, however, is that the psychiatrist is still faced with some practical decisions. In fact, the psychiatrist may share some of the same concerns regarding dissociative disorders, but this conceptual issue may not be particularly relevant in dealing with the patient. Legal requirements or hospital policies might require that the patient be given a traditional diagnosis. The psychiatrist may also have to decide whether to give antipsychotic medication, electroconvulsive therapy, or psychotherapy. An effective examiner should be able to see beyond possible conceptual differences and instead address practical considerations. A psychiatrist may refer a defensive patient who cannot or will not verbalize his or her concerns and ask whether this person has schizophrenia. Beyond this diagnosis are factors such as the quality of the patient's thought processes and whether the person poses a danger to him- or herself or to others. Thus, the effective examiner must translate his or her findings into a conceptual model that is both understandable by a psychiatrist and useful from a task-oriented point of view.

General Medical Setting

It has been estimated that as many as two-thirds of patients seen by physicians have primarily psychosocial difficulties, and of those with clearly established medical diagnoses, between 25% and 50% have psychological disorders in addition to medical

ones (Asaad, 2000; Katon & Walker, 1998; McLeod, Budd, & McClelland, 1997; Mostofsky & Barlow, 2000). Most of these psychological difficulties are neither diagnosed nor referred for treatment (American Journal of Managed Care, 1999; Blount et al., 2007; Borus, Howes, Devins, & Rosenberg, 1988; Mostofsky & Barlow, 2000). In addition, many traditionally “medical” disorders, such as coronary heart disease, asthma, allergies, rheumatoid arthritis, ulcers, and headaches, have been found to possess a significant psychosocial component (Blount et al., 2007; Groth-Marnat & Edkins, 1996). Not only are psychological factors related to disease; of equal importance, they are related to the development and maintenance of health. In addition, the treatment and prevention of psychosocial aspects of “medical” complaints have been demonstrated to be cost-effective for areas such as preparation for surgery, smoking cessation, rehabilitation of chronic pain patients, obesity, interventions for coronary heart disease, and patients who are somatizing psychosocial difficulties (Blount et al., 2007; Chiles, Lambert, & Hatch, 1999; Groth-Marnat & Edkins, 1996; Groth-Marnat, Edkins, & Schumaker, 1995; Sobel, 2000). A complete approach to the patient, then, involves an awareness of the interaction among physical, psychological, and social variables (Kaslow et al., 2007; G. Schwartz, 1982). Thus, psychologists have the potential to make an extremely important contribution. To adequately work in general medical settings, psychologists must become familiar with medical descriptions, which often means learning a complex and extensive vocabulary (see J. D. Robinson & Baker, 2006). Another issue is that, even though they often draw information from several sources to aid in decision making, physicians must take ultimate responsibility for their decisions.

The most frequent situations in which physicians might use the services of a psychologist involve the presence of an underlying psychological disorder, possible emotional factors associated with medical complaints, assessment for neuropsychological deficit, psychological treatment for chronic pain, the treatment of chemical dependency, patient management, and case consultation (Bamgbose et al., 1980; Groth-Marnat, 1988; Pincus, Pechura, Keyser, Bachman, & Houtsinger, 2006). Regardless of whether a medical exam uncovers any physical basis for a patient’s complaints, the physician still has to devise some form of treatment or at least an appropriate referral. This process is crucial in that a significant portion of patients referred to physicians do not have any detectable physical difficulties, and their central complaint is likely to be psychological (Asaad, 2000; Blount et al., 2007; Maruish & Nelson, 2014; Mostofsky & Barlow, 2000). The psychologist can then elaborate and specify how a patient can be treated for possible psychosocial difficulties (Kaslow et al., 2007; Wickramasekera, 1995a, 1995b). Doing this may require using not only the standard assessment instruments, but also more specialized ones, such as the Millon Behavioral Health Inventory or the Millon Behavioral Medicine Diagnostic (Bockian, Meagher, & Millon, 2000; Maruish, 2000; Millon, 1997).

Another area that has greatly increased in importance is the psychological assessment of a patient’s neuropsychological status (see Chapter 12). Whereas physicians attempt to detect physical lesions in the nervous system, the neuropsychologist has traditionally been more concerned with the status of higher cortical functions. Another way of stating this is that physicians evaluate how the *brain* is functioning, while neuropsychologists evaluate how the *person* is functioning as a result of possible

brain abnormalities. The typical areas of assessment focus primarily on the presence of possible intellectual deterioration in areas such as memory, sequencing, abstract reasoning, spatial organization, and executive abilities (Groth-Marnat, 2000b). Such referrals, or at least screening for neuropsychological deficit, typically account for approximately one-third of all psychological referrals in psychiatric and medical settings. In the past, neuropsychologists have been asked to help determine whether a patient's complaints were "functional" or "organic." The focus now is more on whether the person has neuropsychological deficits that may contribute to or account for observed behavioral difficulties than on either/or distinctions (Loenberger, 1989). Physicians often want to know whether a test profile suggests a specific diagnosis, particularly malingering, conversion disorder, hypochondriasis, organic brain syndrome, or depression with pseudoneurological features. Further issues that neuropsychologists often address include the nature and extent of identified lesions, localization of lesions, emotional status of neurologically impaired patients, extent of disability, and suggestions for treatment planning such as recommendations for cognitive rehabilitation, vocational training, and readjustment to family and friends (Lemsky, 2000; Lezak, Howieson, Bigler, & Tranel, 2012; P. J. Snyder, Nussbaum, & Robins, 2006).

A physician might also request a psychologist to conduct a presurgical evaluation to assess the likelihood of a serious stress reaction to surgery. Finally, physicians—particularly pediatricians—are often concerned with detecting early signs of serious psychological disorder, which may have been brought to their attention by parents, other family members, or teachers. In such situations, the psychologist's evaluation should assess not only the patient's current psychological condition but also the contributing factors in his or her environment and should provide a prediction of the patient's status during the next few months or years. When the patient's current condition, current environment, and future prospects have been evaluated, the examiner can then recommend the next phase in the intervention process. A psychologist may also consult with physicians to assist them in effectively discussing the results of an examination with the patient or the patient's family.

Legal Context

During the past 40 years, the use of psychologists in legal settings has become more prevalent, important, and accepted (see Goldstein, 2007; Otto & Heilburn, 2002). Psychologists might be called in at any stage of legal decision making. During the investigation stage, they might be consulted to assess the reliability or quality of information presented by a witness. The prosecuting attorney might also need to have a psychologist evaluate the quality of another mental health professional's report, evaluate the accused person's competency, or help determine the specifics of a crime. A defense attorney might use a psychologist to help in supporting an insanity plea, to help in jury selection, or to document that brain damage has occurred. A judge might use a psychologist's report as one of a number of factors to help determine a sentence, a penal officer might wish consultation to help determine the type of confinement or level of dangerousness, or a parole officer might need assistance to help plan a rehabilitation program. Even though a psychologist might write a legal report, he or she is likely to actually appear in court in only about 1 in every 10 cases.

The increasing use and acceptance of psychologists in legal contexts have resulted in a gradual clarification of their roles (Goldstein, 2007; Otto & Heilburn, 2002) as well as a proliferation of forensic specific assessment instruments (Archer, 2006; Archer, Buffington-Vollum, Stredny, & Handel, 2006; Heilbrun, Marczyk, & Dematteo, 2002). However, acclimatizing to the courtroom environment is often difficult for multiple reasons, including the quite distinct differences between courtrooms and clinics, as well as the need to become familiar with specialized legal terms like *diminished capacity* and *insanity*. In addition, many attorneys are familiar with the same professional literature that psychologists read and may use this information to discredit a psychologist's qualifications, methods of assessment, or conclusions (Ziskin & Faust, 2008). Psychologists are also required to become increasingly sophisticated in their evaluation of possible malingering and deception (see review on kspope.com/assess/malinger.php).

Each psychologist appearing in court must have his or her qualifications approved. Important areas of consideration are the presence of clinical expertise in treating specialty disorders and relevant publication credits. Evaluation of legal work by psychologists indicates they are generally viewed favorably by the courts and may have reached parity with psychiatrists (Sales & Miller, 1994).

As outlined by the American Board of Forensic Psychology (www.abfp.com), the practice of forensic psychology includes training/consultation with legal practitioners, evaluation of populations likely to encounter the legal system, and the translation of relevant technical psychological knowledge into usable information. Psychologists are used most frequently in child custody cases, competency of a person to dispose of property, juvenile commitment, comprehension of Miranda rights, potential for having given a false confession, and personal injury suits in which the psychologist documents the nature and extent of the litigant's suffering or disability (e.g., stress, anxiety, cognitive deficit).

An essential requirement when working in the legal context is for psychologists to modify their language. Many legal terms have exact and specific meanings that, if misunderstood, could lead to extremely negative consequences. Words such as *incompetent*, *insane*, or *reasonable certainty* may vary in different judicial systems or from state to state. Psychologists must familiarize themselves with this terminology and the different nuances involved in its use. Psychologists may also be requested to explain in detail the meaning of their conclusions and how these conclusions were reached. Whereas attorneys rarely question the actual data that psychologists generate, the inferences and generalizability of these inferences are frequently placed under scrutiny or even attacked. Often this questioning can seem rude or downright hostile, but in most cases, attorneys are merely doing their best to defend their client. Proper legal protocol also requires that the psychologist answer questions directly rather than respond to the implications or underlying direction suggested by the questions. Furthermore, attorneys (or members of the jury) may not be trained in or appreciate the scientific method, which is the mainstay of a psychologist's background. In contrast, attorneys are trained in legal analysis and reasoning, which subjectively focus on the uniqueness of each case rather than on a comparison of the person to a statistically relevant normative group (see Hilsenroth & Stricker, 2004).

Two potentially problematic areas lie in evaluating insanity and evaluating competency. Although the insanity plea has received considerable publicity, very few people

make the appeal; and, of those who do, few have it granted. It is usually difficult for an expert witness to evaluate such cases because of the problem of possible malingering to receive a lighter sentence and the possible ambiguity of the term *insanity*. Usually a person is considered insane in accordance with the McNaughton Rule, which states that persons are not responsible if they did not know the nature and extent of their actions and if they cannot distinguish that what they did was wrong according to social norms. In some states, the ambiguity of the term is increased because defendants can be granted the insanity plea if it can be shown they were insane at the time of the incident. Other states include the clause of an “irresistible impulse” to the definition of insanity. Related to insanity is whether the defendant is competent to stand trial. *Competence* is usually defined as the person’s ability to cooperate in a meaningful way with the attorney, understand the purpose of the proceedings, and understand the implications of the possible penalties. To increase the reliability and validity of competency and insanity evaluations, specialized assessment techniques have been developed; these include the MacArthur Competence Assessment Tool (Poythress et al., 1999), the Evaluation of Competency to Stand Trial–Revised (R. Rogers, Tillbrook, & Sewell, 2004), and the Rogers Criminal Responsibility Assessment Scales (R. Rogers, 1984).

The prediction of dangerousness has also been a problematic area. Because actual violent or self-destructive behavior is a relatively unusual (low base rate) behavior, any cutoff criteria will typically produce a high number of false positives (Mulvey & Cauffman, 2001). Thus, people incorrectly identified may potentially be detained and will understandably be upset. However, the negative result of failure to identify and take action against people who are potentially violent makes erring on the side of caution more acceptable. Attempts to use special scales on the Minnesota Multiphasic Personality Inventory (MMPI; Overcontrolled Hostility Scale; Megargee & Mendelsohn, 1962) or a 4-3 code type (see Chapter 7) have not been found to be sufficiently accurate for individual decision making. However, significant improvements have been made in predicting dangerousness and reoffending by assessing for the presence of antisocial features and using actuarial strategies, collateral sources, formal ratings, and summed ratings, which include relevant information on developmental influences, possible events that lower thresholds, arrest record, life situation, and situational triggers, such as interpersonal stress and substance intoxication (Monahan & Steadman, 2001; Monahan et al., 2000; Tolman & Rotzien, 2007). The legal/justice system is most likely to give weight to those individual assessment strategies that combine recidivism statistics, tests specifically designed to predict dangerousness, summed ratings, and double administrations of psychological tests to assess change over time. Frequently used tests include the Historical Clinical Risk–20 (for violence risk assessment; Webster, Douglas, Eaves, & Hart, 1997) and the Static 99 (for sexual reoffending risk; Hanson & Thornton, 1999). In contrast, informal clinical interviews are clearly considered to be insufficient (Tolman & Rotzien, 2007).

Psychologists are sometimes asked to help with child custody decisions. Guidelines for developing child custody evaluations and child protection evaluations have been developed by the American Psychological Association: (Guidelines for Child Custody Evaluations in Family Law Proceedings, 2010; www.apa.org/practice/guidelines/child-custody.aspx). The central consideration is to determine which arrangement is in the child’s best interest. Areas to be considered include the mental health of the

parent, the quality of love and affection between the parent and child, the nature of the parent-child relationship, and the long-term potential effects of the different decisions on the child (M. J. Ackerman, 2006a, 2006b). Often psychological evaluations are conducted on each member of the family using traditional testing instruments. Specific tests, such as the Bricklin Perceptual Scales (Bricklin, 1984), have also been developed.

A final, frequently requested service is to aid in the classification of inmates in correctional settings. One basic distinction is between merely managing the person and attempting a program of rehabilitation. Important management considerations are levels of suicide risk, appropriateness of dormitory versus a shared room, possible harassment from other inmates, or degree of dangerousness to others. Rehabilitation recommendations may need to consider the person's educational level, interests, skills, abilities, and personality characteristics related to employment.

Academic/Educational Context

Psychologists are frequently called on to assess children who are having difficulty, or may need special placement, in the school system. The most important areas are evaluating the nature and extent of a child's learning difficulties, measuring intellectual strengths and weaknesses, assessing behavioral difficulties, creating an educational plan, estimating a child's responsiveness to intervention, and recommending changes in a child's program or placement (Sattler, 2008, 2014). Any educational plan should be sensitive to the interactions among a child's abilities, diversity considerations, the child's personality, the characteristics of the teacher, and the needs and expectations of the parents.

A typical educational placement begins with a visit to the classroom for observation of a child's behavior under natural conditions. A valuable aspect of this visit is to observe the interaction between the teacher and child. Often, behavioral difficulty is closely linked with the child-teacher interaction. Sometimes the teacher's style of responding to a student can be as much a part of the problem as the student. Consequently, classroom observations can cause discomfort to teachers and should be handled sensitively.

Observing the child in a wider context is, in many ways, contrary to the tradition of individual testing. However, individual testing all too frequently provides a relatively limited and narrow range of information, especially because children are not reliable self-reporters and parents or caregivers may be biased. If testing is combined with a family or classroom assessment, additional crucial data may be collected, though there may also be significant resistance. This resistance may result from legal or ethical restrictions regarding the scope of the services the school can provide or the demands that a psychologist can make on the student's parents. Often there is an initial focus on, and need to perceive, the student as a "problem child" or "identified patient." This focus may obscure larger, more complex, and yet more significant, issues, such as marital conflict, a disturbed teacher, misunderstandings between teacher and parents, or a conflict between the school and the parents. All or some of the involved individuals may have an investment in perceiving the student as the person with the problem, rather than acknowledging that a disordered school system or family difficulties may be responsible. An individually oriented assessment may be conducted with excellent

interpretations, but unless wider contexts are considered, understood, and addressed, the assessment may very well be ineffective in solving both the individual difficulties and the larger organizational or interpersonal problems.

Most assessments of children in a school context include behavioral observations, a test of intellectual abilities such as the Weschsler Intelligence Scale for Children–V, Stanford Binet–V, Woodcock-Johnson Psychoeducational Battery–IV (Woodcock, Schrank, Mather, & McGrew, 2014), or Kaufman Assessment Battery for Children–II (K-ABC-II; Kaufman & Kaufman, 2004), and tests of emotional and behavioral functioning. In the past, assessment of children’s emotional functioning generally relied on projective techniques. However, many projective tests have been found to have inadequate psychometric properties and are time consuming to administer, score, and interpret. As a result, a wide variety of behavioral ratings instruments have begun to replace the use of projective instruments (Kamphaus, Petoskey, & Rowe, 2000). These include the Achenbach Child Behavior Checklist (Achenbach & Rescorla, 2001), Conners–3 Parent and Teacher Rating Scales (Conners, 2008), and the Behavior Assessment System for Children–3 (BASC-3; C. R. Reynolds & Kamphaus, 2015). A number of sound objective instruments, such as the Personality Inventory for Children–2 (PIC-2; Lachar & Gruber, 2001), have also been developed. This inventory was designed along similar lines as the MMPI but is completed by a child’s parent. It produces four validity scales to detect faking and 12 clinical scales, such as Depression, Family Relations, Delinquency, Anxiety, and Hyperactivity. Assessment of adolescent personality can be done effectively with the MMPI-A or the Millon Adolescent Clinical Inventory (MACI; Millon, 1993). Additional well-designed scales that are increasingly used are the Vineland Adaptive Behavior Scales–II (Sparrow, Cicchetti, & Balla, 2005), the Wechsler Individual Achievement Test–III (WIAT-III; Pearson, 2009a), and the Wide Range Achievement Test–IV (WRAT-IV; Wilkinson & Robertson, 2007).

Any report written for an educational setting should focus not only on a child’s weaknesses but also on his or her strengths. Understanding of a child’s strengths can potentially be used to increase a child’s self-esteem as well as to create change in a wide context. Recommendations should be realistic and practical. Recommendations can be developed most effectively when the clinician has a thorough understanding of relevant resources in the community, the school system, and the classroom environment. This understanding is particularly important because the quality and resources available in one school or school system can vary tremendously from another. Recommendations typically specify which skills need to be learned, how these can be learned, a hierarchy of objectives, and possible techniques for reducing behaviors that make learning difficult.

Recommendations for special education should be made only when a regular class would clearly not be equally beneficial. However, the recommendations are not the end product. They are beginning points that should be elaborated and modified depending on the initial results. Ideally, a psychological report should be followed up with continuous monitoring.

The psychoeducational assessment of children should be carried out in two phases. The first phase should assess the nature and quality of the child’s learning environment. If the child is not exposed to adequate quality instruction, he or she cannot be

expected to perform well. Thus, it must first be demonstrated that a child is struggling despite appropriate instruction. The second phase involves a comprehensive assessment battery, which includes measures of intellectual abilities, academic skills, adaptive behavior, and screening out any biomedical disorders that might disrupt learning. Intellectual abilities might involve memory, spatial organization, abstract reasoning, and sequencing. Regardless of students' academic and intellectual abilities, they will not perform well unless they have relevant adaptive abilities, such as social skills, adequate motivation and attention, and ability to control impulses. Assessing a child's values and attitudes toward education may be particularly important because they determine whether the student is willing to use whatever resources he or she may have. Likewise, the person's level of personal efficacy helps to determine whether he or she is able to perform behaviors leading toward attaining the goals the person values. Physical difficulties that might interfere with learning include poor vision, poor hearing, hunger, extreme fatigue, malnutrition, or endocrine dysfunction.

The preceding considerations clearly place the assessment of children in educational settings into a far wider context than merely the interpretation of test scores. Relationships among the teacher, family, and student need to be assessed, along with the relative quality of the learning environment. Furthermore, the child's values, motivation, and sense of personal efficacy need to be taken into consideration, along with possible biomedical difficulties. Examiners need to become knowledgeable regarding the school and community resources as well as learn population-specific instruments that have demonstrated relatively high levels of reliability and validity.

Psychological Clinic

In contrast to the medical, legal, and educational institutions where the psychologist typically serves as a consultant to the decision maker, the psychologist working in a psychological clinic often is the decision maker. A number of frequent types of referrals come into the psychological clinic. Perhaps the most common ones are individuals who are self-referred and are seeking relief from psychological turmoil. For many of these individuals, extensive psychological testing may not be relevant and, in fact, may be contraindicated, as their diagnoses and issues may be relatively straightforward and time spent in testing could best be applied toward treatment. However, brief instruments targeted toward assessing client characteristics most relevant to treatment planning can help develop treatments that will speed the rate of improvement as well as optimize outcome (see Chapters 13 and 14). Brief instruments can also be used to monitor response to therapy or inform relevant alterations, thus increasing the likelihood of successful intervention (Lambert & Hawkins, 2004). In addition, there may be certain groups of self-referred clients about whom the psychologist may question whether the treatment available in a psychological clinic is appropriate. These clients can include persons with extensive medical problems, individuals with legal complications that need additional clarification, and persons who may require higher levels of care. With these cases, it might be necessary to obtain additional information through psychological testing. The main purpose of the testing would be to aid in decision making rather than to serve as a direct source of help for the client. Still other clients in clinics who may benefit from psychological testing are those who are being seen in the clinic

already, either who have unclear diagnoses or whose treatment has stalled or plateaued. These cases may benefit from the clear guidance of a comprehensive assessment.

Two other situations in which psychological assessment may be warranted involve children who are referred by their parents for school or behavioral problems and referrals from other decision makers. When referrals are made for poor school performance or involving legal complications, special precautions must be taken before testing. Primarily, the clinician must develop a complete understanding of the client's social network and the basis for the referral. This complete understanding may include a history of previous attempts at treatment and a summary of the relationship among the parents, school, courts, and child. Usually a referral comes at the end of a long sequence of events, and it is important to obtain information regarding these events. After the basis of the referral has been clarified, the clinician may decide to have a meeting with different individuals who have become involved in the case, such as the school principal, previous therapists, probation officer, attorney, or teacher. This meeting may uncover myriad issues that require decisions, such as referral for family therapy, placement in special education, a change in custody agreements between divorced parents, individual therapy of other members of the family, and a change in school. All of these issues may affect the relevance of, and approach to, testing, but they may not be apparent if the initial referral question is taken at face value. Sometimes psychologists are also confronted with referrals from other decision makers. For example, an attorney may want to know if an individual is competent to stand trial. Other referrals may involve a physician who wants to know whether a patient with a head injury can readjust to his or her work environment or drive a car, or the physician may need to document changes in a patient's recovery.

So far, this discussion on the different settings in which psychological testing is used has focused on when to test and how to clarify the manner in which tests can be most helpful in making decisions. Several additional summary points must be stressed. As has been discussed previously, a referral source sometimes is unable to adequately formulate the referral question. In fact, the referral question is usually neither clear nor concise. It is the evaluator's responsibility to look beyond the referral question and determine the basis for the referral in its widest scope. Thus, psychologists must develop an understanding of the complexity of the client's social setting, including interpersonal factors, family dynamics, and the sequence of events leading to the referral. In addition to clarifying the referral question, a second major point is that psychologists are responsible for developing knowledge about the setting for which they are writing their reports. This knowledge includes learning the proper language, the roles of the individuals working in the setting, the choices facing decision makers, and the philosophical and theoretical beliefs they adhere to. It is also important that clinicians understand the values underlying the setting and assess whether these values coincide with their own. For example, psychologists who do not believe in aversion therapy, capital punishment, or electroconvulsive therapy may come into conflict while working in certain settings. Psychologists, thus, should clearly understand how the information they give their referral source will be used. It is essential for them to appreciate that they have a significant responsibility, because decisions made regarding their clients, which are often based on assessment results, can frequently be major changing points in a client's life. If the possibility exists for the information to be used in a manner

that conflicts with the evaluator's value system, he or she should reconsider, clarify, or possibly change his or her relationship to the referral setting.

A final point is that clinicians should not allow themselves to be placed into the role of a "testing technician" or psychometrist. This role ultimately does a disservice to the client, the practitioner, and the profession. Clinicians should not merely administer, score, and interpret tests but should also understand the total referral context in its broadest sense. This means they also take on the role of an expert who can integrate data from a variety of sources. Tests, by themselves, are limited in that they are not flexible or sophisticated enough to address themselves to complex referral questions. D. Levine (1981) wrote:

[The formal research on test validity is] not immediately relevant to the practical use of psychological tests. The question of the value of tests becomes not "Does this test correlate with a criterion?" or "Does the test accord with a nomological net?" but rather "Does the use of the test improve the success of the decision making process?" by making it either more efficient, less costly, more accurate, more rational, or more relevant. (p. 292)

All of these concerns are consistent with the emphasis on an evaluator fulfilling the role of an expert clinician performing psychological assessment rather than a psychometrist acting as a technician.

ETHICAL PRACTICE OF ASSESSMENT

Ethical guidelines reflect values that professional psychology endorses. These values include client safety, confidentiality, the reduction of suffering, fairness, and advancing science. These guidelines have largely evolved through careful consideration of how these values are expressed in ideal practice. Notably, many of the ethical codes have been refined due to conflicts and criticisms related to assessment procedures. Criticism has been directed at the use of tests in inappropriate contexts, confidentiality, cultural bias, invasion of privacy, release of test data, and the continued use of tests that are inadequately validated. These criticisms have resulted in restrictions on the use of certain tests, greater clarification within the profession regarding ethical standards, and, unfortunately, increased skepticism from the public. To deal with these potential difficulties as well as conduct useful and accurate assessments, clinicians need to be aware of the ethical use of assessment tools. The American Educational Research Association (AERA) and other professional groups have published guidelines for examiners in their *Standards for Educational and Psychological Tests* (1999) and the *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 2002). A special series in the *Journal of Personality Assessment* (Russ, 2001) also elaborated on ethical dilemmas found in training, medical, school, and forensic settings. The next section outlines the most important of these guidelines along with additional related issues.

Developing a Professional Relationship

Assessment should be conducted only in the context of a clearly defined professional relationship. This means that the nature, purpose, and conditions of the relationship

must be discussed and agreed on. Usually the clinician provides relevant information, followed by the client's signed consent. Information conveyed to the client usually relates to the type and length of assessment, alternative procedures, details related to appointments, the nature and limits of confidentiality, financial requirements, and additional general information that might be relevant to the unique context of an assessment (see Pope, 2007a, 2007b, and Zuckerman, 2003, for specific guidelines, formats, and forms for informed consent).

An important area to be aware of is the impact the quality of the relationship can have on both assessment results and the overall working relationship. It is the examiner's responsibility to recognize the possible influences he or she may exert on the client and to optimize the level of rapport. For example, enhanced rapport with older children (but not younger ones) involving verbal reinforcement and friendly conversation has been shown to increase WISC-R scores by an average of 13 IQ points compared with an administration involving more neutral interactions (Feldman & Sullivan, 1971). This is a difference of nearly 1 full standard deviation. It has also been found that mildly disapproving comments, such as "I thought you could do better than that," resulted in significantly lowered performance when compared with either neutral or approving comments (Witmer, Bornstein, & Dunham, 1971). In a review of 22 studies, Fuchs and Fuchs (1986) concluded that, on average, IQ scores were 4 points higher when the examiner was familiar with the child being examined than when he or she was unfamiliar with the child. This trend was particularly pronounced for children from lower socioeconomic status. Whereas there is little evidence (Lefkowitz & Fraser, 1980; Sattler, 1973a, 1973b; Sattler & Gwynne, 1982) to support the belief that African American students have lower performance when tested by European American examiners, it has been suggested that African American students are more responsive to tangible reinforcers (money, candy) than are European American students, who generally respond better to verbal reinforcement (Schultz & Sherman, 1976). However, in a later study, Terrell, Taylor, and Terrell (1978) demonstrated that the main factor was the cultural relevance of the response. They found a remarkable 17.6-point increase in IQ scores when African American students were encouraged by African American examiners with culturally relevant comments such as "nice job, blood" or "good work, little brother." Thus, positive rapport and feedback, especially if that feedback is culturally relevant, can significantly improve test performance. As a result, the feedback, and level of rapport should, as much as possible, be held constant from one test administration to the next.

A variable extensively investigated by Rosenthal and his colleagues is that a researcher/examiner's expectations can influence another person's level of performance (R. Rosenthal, 1966). This finding has been demonstrated with humans as well as with laboratory rats. For example, when an experimenter was told to expect better performances from rats that were randomly selected from the same litter as "maze bright" (compared with "maze dull"), the descriptions of the rats' performance given by the experimenter conformed to the experimenter's expectations (R. Rosenthal & Fode, 1963). Despite criticisms that have been leveled at his studies and the finding that the magnitude of the effect was not as large as originally believed (Barber & Silver, 1968; Elashoff & Snow, 1971), Rosenthal maintains that an expectancy effect exists in some situations and suggests that the mechanisms include minute, nonverbal

behaviors (H. Cooper & Rosenthal, 1980). He maintains that the typical effects on an individual's performance are usually small and subtle and occur in some situations but not in others. The obvious implication for clinicians is that they should continually question themselves regarding their expectations of clients and check to see whether they may in some way be communicating these expectations to their clients in a manner that confounds the results.

An additional factor that may affect the nature of the relationship between the client and the examiner is the client's relative emotional state. It is particularly important to assess the degree of the client's motivation and his or her overall level of anxiety. There may be times in which it would be advisable to discontinue testing because situational emotional states may significantly influence the results of the tests. At the very least, examiners should consider the possible effects of emotional factors and incorporate these into their interpretations. For example, it might be necessary to increase the estimate of a client's optimal intellectual functioning if the client were obviously extremely anxious during administration of an intelligence test.

A final consideration, which can potentially confound both the administration and, more commonly, the scoring of responses, is the degree to which the examiner likes the client and perceives him or her as warm and friendly. Several studies (Sattler, Hillix, & Neher, 1970; Sattler & Winget, 1970) have indicated that the more the examiner likes the client, the more likely he or she will be to score an ambiguous response in a direction favorable to the client. Higher scores can occur even on items in which the responses are not ambiguous (Egeland, 1969; Simon, 1969). Thus, "hard" scoring, as opposed to more lenient scoring, can occur at least in part because of the degree of subjective liking the examiner feels toward the client. Again, examiners should continually check themselves to assess whether their relationship with the client is interfering with the objectivity and standardization of the test administration and scoring.

Issues Related to Informed Consent

Psychologists should obtain informed consent for assessment procedures. Any consent involves a clear explanation of what procedures will occur, the relevance of the testing, and how the results will be used (see Pope, 2007a; kspope.com/consent/index.php). This means that examiners should always have a clear conception of the specific reasons for giving a test. It should be stressed what information is confidential and what limitations to confidentiality exist. Exceptions to confidentiality may occur in situations involving child/elder abuse, danger to self or others, and information that has been requested based on a subpoena. The information should be provided in clear, straightforward language that can be understood by the client. Unfortunately, many formal consent forms are written at a level far above the reading comprehension level of a large proportion of clients.

Informed consent involves communicating not only the rationale for testing but also the kinds of data obtained and the possible uses of the data, when this will not adversely affect the results. This fact does not mean the client should be shown the specific test subscales beforehand, but rather that the nature and intent of the test should be described in a general way. For example, if a client is told that a scale measures "sociability," this foreknowledge might alter the test's validity in that the client

may answer questions based on popular, but quite possibly erroneous, stereotypes. Introducing the test format and intent in a simple, respectful, and forthright manner significantly reduces the chance that the client will perceive the testing situation as an invasion of privacy.

Sometimes clinicians will have provided clear information and the client will have agreed to the procedures, but unforeseen events not covered in the information may occur. This might happen when the examiner discovers aspects of the client that the client would rather keep secret. Thus, assessment may entail an invasion of privacy. The Office of Science and Technology (1967), in a report entitled "Privacy and Behavioral Research," defined privacy as "the right of the individual to decide for him/herself how much he will share with others his thoughts, feelings, and facts of his personal life." This right is considered to be "essential to insure dignity and freedom of self determination" (p. 2). The invasion of privacy issue usually becomes most controversial with personality tests, as items relating to motivational, emotional, and attitudinal traits are often disguised. Thus, persons may unknowingly reveal characteristics about themselves that they would rather keep private. Similarly, many individuals consider their IQ scores to be highly personal. Public concern over this issue culminated in an investigation by the Senate Subcommittee on Constitutional Rights and the House Subcommittee on Invasion of Privacy. Neither of these investigations found evidence of deliberate or widespread misuse of psychological tests (Brayfield, 1965).

Dahlstrom (1969) argued that public concern over the invasion of privacy is based on two basic issues. The first is that tests have been oversold to the public, with a resulting exaggeration of their scope and accuracy. The public is usually not aware of the limitations of test data and may often feel that tests are more capable of discovering hidden information than they actually are. The second misconception is that it is not necessarily wrong to obtain information about persons that they either are unaware of themselves or would rather keep private. The more important issue is how the information is used. Furthermore, the person who controls where and how this information is used is generally the client. The ethical code of the American Psychological Association (2002) specifically states that information derived by a psychologist from any source can be released only with the permission of the client. Although there may be exceptions regarding the rights of minors or when clients are a danger to themselves or others, the ability to control the information is usually clearly defined as being held by the client. Thus, the public can be uneducated regarding its rights and can underestimate the power it has in determining how the test data will be used.

Whereas concerns about invasion of privacy relate to the discovery and misuse of information that clients would rather keep secret, *inviolacy* involves the actual negative feelings created when clients are confronted with the test or test situation. Inviolacy is particularly relevant when clients are asked to discuss information they would rather not think about. For example, the MMPI contains questions about many ordinarily taboo topics relating to sexual practices, toilet behavior, bodily functions, and personal beliefs about human nature. Such questions may produce anxiety by making examinees think about deviant thoughts or repressed unpleasant memories. Many individuals obtain a certain degree of security and comfort by staying within familiar realms of thought. Even to be asked questions that may indicate the existence of unusual alternatives can serve as an anxiety-provoking challenge to personal rules

and norms. This problem is somewhat related to the issue of invasion of privacy, and it too requires one-to-one sensitivity as well as clear and accurate disclosure about the assessment procedure.

Another issue is that during personnel evaluations, participants might feel pressured to reveal personal information on tests because they aspire to a certain position. Also, applicants may unknowingly reveal information because of subtle, nonobvious test questions, and, perhaps more importantly, they have no control over the inferences that examiners make about the test data. However, if a position requires careful screening and if serious negative consequences may result from poor selection, it is necessary to evaluate an individual as closely as possible. Thus, careful testing may be required to select personnel in the police, in delicate military positions, or for important public duty overseas.

In a clinical setting, obtaining personal information regarding clients usually does not present problems. The agreement that the information be used to help clients develop new insights and change their behavior is generally clear and straightforward. However, should legal difficulties arise relating to areas such as child abuse, involuntary confinement, or situations in which clients may be a danger to themselves or others, ethical questions often arise. Usually there are general guidelines regarding the manner and extent to which information should be disclosed. These are included in the American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* (2002), and test users are encouraged to familiarize themselves with these guidelines. Professional psychologists can also consult with colleagues, their insurance companies, or the APA's ethics office (apa.org/ethics).

Labeling and Restriction of Freedom

When individuals are given a medical diagnosis for physical ailments, the social stigmas are usually relatively mild. In contrast are the potentially damaging consequences of many psychiatric diagnoses. A major danger is the possibility of creating a self-fulfilling prophecy based on the expected roles associated with a specific label. Many of these expectations are communicated nonverbally and are typically beyond a person's immediate awareness (H. Cooper & Rosenthal, 1980; R. Rosenthal, 1966). Other self-fulfilling prophecies may be less subtle; for example, a juvenile with minor but poor sexual boundaries might be labeled as a "sex offender," which would then result in quite restrictive treatment along with quite public distribution of the label.

Just as labels imposed by others can have negative consequences, self-acceptance of labels can likewise be detrimental. Clients may use their labels to excuse or deny responsibility for their behavior. This is congruent with the medical model, which usually assumes that a "sick" person is the victim of an "invading disorder." Thus, in our society, "sick" persons are not considered responsible for their disorders. However, the acceptance of this model for behavioral problems may perpetuate behavioral disorders because persons see themselves as helpless, passive victims under the power of mental health "helpers" (Szasz, 1987). This sense of helplessness may serve to lower people's ability to deal effectively with new stress. In contrast to this sense of helplessness is the belief that clients require an increased sense of responsibility for their lives and actions to effectively change their behavior.

A final difficulty associated with labeling is that it may unnecessarily impose limitations on either an individual or a system by restricting progress and creativity. For example, an organization may conduct a study to determine the type of person who has been successful at a particular type of job and may then develop future selection criteria based on this study. This can result in the future selection of relatively homogeneous employees, which in turn could prevent the organization from changing and progressing. There may be a narrowing of the “talent pool,” in which people with new and different ideas are never given a chance. In other words, what has been labeled as adaptive in the past may not be adaptive in the future. One alternative to this predicament is to look at future trends and develop selection criteria based on these trends. Furthermore, diversity might be incorporated into an organization so that different but compatible types can be selected to work on similar projects. Thus, clinicians should be sensitive to the potential negative impact resulting from labeling by outside sources or by self-labeling, as well as to the possible limiting effects that labeling might have.

Competent Use of Assessment Instruments

To correctly administer and interpret psychological tests, an examiner must have proper training, which generally includes adequate graduate coursework combined with lengthy supervised experience (Turner, DeMers, Fox, & Reed, 2001). Clinicians should have knowledge of tests and test limitations and should be willing to accept responsibility for competent test use. Intensive training is particularly important for individually administered intelligence tests and for the majority of personality tests. Students who are administering tests as part of a class requirement are generally not yet adequately trained to administer and interpret tests professionally. Thus, test results obtained by students have questionable validity, and students should clearly inform their subjects that the purpose of their testing is for training purposes only.

In addition to the preceding general guidelines for training, examiners should also acquire a number of specific skills (Moreland, Eyde, Robertson, Primoff, & Most, 1995; Turner et al., 2001). These include the ability to evaluate the technical strengths and limitations of a test, the selection of appropriate tests, knowledge of issues relating to the test’s reliability and validity, and interpretation with diverse populations. Examiners need to be aware of the material in the test manual as well as relevant research both on the variable the test is measuring and the status of the test since its publication. This is particularly important with regard to newly developed subgroup norms and possible changes in the meaning of scales resulting from further research. After examiners evaluate a test itself, they must also be able to evaluate whether the purpose and context for which they would like to use it are appropriate. Sometimes an otherwise valid test can be used for purposes it was not intended for, resulting in either invalid or useless inferences based on the test data. In addition, examiners must be continually aware of, and sensitive to, conditions affecting the examinee’s performance. These conditions may include expectations on the part of the examiner, minor variations from the standardized instructions, degree of rapport, mood of the examinee, or timing of the test administration in relation to an examinee’s life changes. To help develop accurate conclusions, examiners should have a general knowledge of the diversity of human behavior. Different considerations and interpretive strategies may be necessary

for various ethnic groups, sexes, sexual orientations, or persons from different countries (see Dana, 2005; Nguyen, Huang, Arganza, & Liao, 2007). A final consideration is that, if interns or technicians are administering the tests, an adequately trained psychologist should be available as a consultant or supervisor.

Specific data-based guidelines for test user qualifications have been developed by relevant professional organizations (American Psychological Association, 1987; Turner et al., 2001), and these guidelines have been incorporated by most organizations selling psychological tests. Qualification forms request information regarding the purpose for using tests (counseling, research, personnel selection), area of professional expertise (marriage and family, social work, school), level of training (degrees, licenses), specific courses taken (descriptive statistics, career assessment), and quality control over test use (test security, appropriate tailoring of interpretations). Persons completing the forms certify that they possess appropriate training and competencies and agree to adhere to ethical guidelines and legal regulations regarding test use.

In addition to being appropriately trained to use tests themselves, psychologists should not promote the use of psychological techniques by persons who are not qualified. This does not mean that all psychological tests should be used exclusively by psychologists, as many tests are available to other professionals. However, psychologists should be aware of which tests require a high level of training (e.g., individually administered IQ tests) and those that are more generally available.

One of the important aspects of competent test use is that tests should be used only for the purposes they were designed for. Typically, tests being extended beyond what they were designed for have been done in good faith and with good intentions. For example, an examiner might use a Thematic Apperception Test or Rorschach as the primary means of inferring an individual's IQ. Similarly, the MMPI-2 or MCMI-IV, which were designed to assess the extent of psychopathology in an individual, might be inappropriately used to assess a normal person's level of functioning. Although some conclusions can be drawn from the MMPI-2 relating to certain aspects of a normal person's functioning, and although IQ estimates based on projectives can be made, they should be considered extremely tentative. These tests were not designed for these purposes, and, as a result, such inferences do not represent the strengths of the tests. A somewhat more serious misuse can occur when a test such as the MMPI-2 is used to screen applicants for some types of personnel selection. Results from MMPI-2-type tests are likely to be irrelevant for assessing most job-related skills. Of equal importance is that the information derived from the MMPI-2 is typically of a highly personal nature and, if used in many types of personnel selection, is likely to represent an invasion of privacy.

Interpretation and Use of Test Results

Interpreting test results should never be considered a simple, mechanical procedure. Accurate interpretation means not simply using norms and cutoff scores but also taking into consideration unique characteristics of the person combined with relevant aspects of the test itself. Whereas tests themselves can be validated, the integration of information from a test battery is far more difficult to validate. It is not infrequent, for example, to have contradictions among different sources of data. It is up to the clinician

to evaluate these contradictions to develop the most appropriate, accurate, and useful interpretations. If the clinician has significant reservations regarding the test interpretation, these should be communicated, usually in the psychological report itself.

A further issue is that test norms and stimulus materials eventually become outdated. As a result, interpretations based on these tests may become inaccurate. For this reason, clinicians need to stay current on emerging research and new versions of tests. A rule of thumb is that if a clinician has not updated his or her test knowledge in at most the past 10 years, he or she is probably not practicing competently.

Part of remaining current means that psychologists should select their testing instruments, as well as any scoring and interpretation services, based on evidence related to the validity of the programs or tests. Part of this selection process requires knowledge of the context of the situation (Turner et al., 2001). A well-validated test might have been found to be quite valid in one context or population but not for another. Another issue that might have ethical considerations is conversion to or use of computerized or Internet-assisted technology (McMinn, Bearse, Heyne, Smithberger, & Erb, 2011; McMinn, Buchanan, Ellens, & Ryan, 1999; McMinn, Ellens, & Soref, 1999). Ultimately, any interpretations and recommendations regarding a client are the responsibility of the clinician. Placing a signature on a report means that the clinician is taking responsibility for the content of the report. Indeed, an important difference between an actuarial formula or automated report and a practitioner is that the practitioner ultimately will be held accountable.

Communicating Test Results

Psychologists should ordinarily give feedback to the client and referral source regarding the results of assessment (Lewak & Hogan, 2003; see Pope, 1992, 2007b, and on kspope.com/assess/feedabs1.php for forms and guidelines). This feedback should be given using clear, everyday language. If the psychologist is not the person giving the feedback, this should be agreed on in advance, and the psychologist should ensure that the person providing the feedback presents the information in a clear, competent manner. Unless the results are communicated effectively, the purpose of the assessment is not likely to be achieved. Effective feedback involves understanding the needs and vocabulary of the referral source, client, and other persons, such as parents or teachers, who may be affected by the test results. Initially, there should be a clear explanation of the rationale for testing and the nature of the tests being administered. This explanation may include the general type of conclusions that are drawn, the limitations of tests, and common misconceptions surrounding the tests or test variables. If a child is being tested in an educational setting, a meeting should be arranged with the school psychologist, parents, teacher, and other relevant persons. Such an approach is crucial for IQ tests, which are more likely to be misinterpreted than achievement tests. Assessment results feedback should be given in terms that are clear and understandable to the receiver. Descriptions are generally most meaningful when performance levels are clearly indicated along with behavioral references. For example, in giving IQ results to parents, it is only minimally relevant to say that their child has an IQ of 130 with relative strengths in spatial organization, even though this may be appropriate language for a formal psychological evaluation. A more effective description might be that their

child is “currently functioning in the top 2% when compared with his or her peers and is particularly good at organizing nonverbal material, such as piecing together puzzles, putting together a bicycle, or building a playhouse.”

In providing effective feedback, the clinician should also consider the personal characteristics of the receiver, such as his or her general educational level, relative knowledge regarding psychological testing, and possible emotional response to the information (Finn, 2007). The emotional reaction is especially important when a client is learning about his or her personal strengths or shortcomings. Facilities should be available for additional counseling, if needed. If properly given, feedback is not merely informative but can actually serve to reduce symptomatic distress and enhance self-esteem (Armengol, Moes, Penney, & Sapienza, 2001; Finn & Tonsager, 1992; Lewak & Hogan, 2003). Thus, providing feedback can actually be part of the intervention process itself. Because psychological assessment is often requested as an aid in making important life decisions, the potential impact of the information should not be underestimated. Clinicians are usually in positions of power, and with that power comes responsibility, as the information clients receive and the decisions they make based on this information are often with them for many years.

Maintenance of Test Security and Release of Test Data

If test materials were widely available, it would be easy for persons to review the tests, learn the answers, and respond according to the impression they would like to make. Thus, the materials would lose their validity. Not only is maintaining test security an ethical obligation, but it is a legal requirement related to trade secrets and agreements made with test publishers when materials are purchased. Psychologists should make all reasonable efforts to ensure that test materials are secure. Specifically, all tests should be kept locked in a secure place, and no untrained persons should be allowed to review them. Any copyrighted material should not be duplicated (see Zuckerman, 2003, for forms and guidelines).

The security of assessment results should also be maintained. This security usually means that only persons designated by the client (often the referral source and client) should see the results. In reality, however, this ethical principle may sometimes be difficult to achieve. For example, many medical contexts expect most relevant treatment information (including psychological assessment results) to be kept in clients' charts. Typically, all members of the treatment team have access to the charts (Claassen & Lovitt, 2001). On one level, this access represents a conflict between psychological and medical guidelines. On another level, it represents a conflict between benefit to the patient (that may be enhanced by the treatment team having access to his or her records) and patient autonomy (patient control over to whom and where information should go). Security of assessment results can also be compromised when a large number of organizations (e.g., insurance company, interacting rehabilitation provider, referral source) all want access to patient records. This issue arises frequently in the managed care environment. The security of client records also becomes more tenuous when large interconnected databases potentially have access to patient data (McMinn, Bearse et al., 2011; McMinn, Buchanan, et al., 1999; McMinn, Ellens et al., 1999).

In some clinical and legal contexts, the court or the opposing counsel may wish to see actual client data. These data can be released if the client authorizes it or if the material has been subpoenaed. Ideally, however, the examiner should recommend that a qualified person be present to explain the results. This recommendation is consistent with the principle that the examiner protect the client from potential harm. If the examiner feels that releasing the test data may result in “substantial harm” to the client or “misuse or misrepresentation of the data” (American Psychological Association, 2002, p. 12), he or she may have the option of refusing to release the data. This situation may result in a conflict between legal and ethical requirements.

One important distinction is between “test data” and “test materials.” The term *test data* refers to raw and scaled scores, such as subscale scores and test profiles. In contrast, *test materials* refers to “manuals, instruments, protocols, and test questions or stimuli” (American Psychological Association, 2002, p. 13). Interestingly, test materials turn into test data when a psychologist places the client’s name on the materials. Since actual items should not be released, it is important for clinicians to make sure they do not place client-identifying information on what might be copyrighted or restricted material. This is crucial since psychologists can release test data, but they cannot release test materials (e.g., actual test items). As stated, the release of test materials would constitute a breach of trade secrets, copyright, and the conditions of purchase (Behnke, 2004). One exception to this point is that the material may be released to persons who are properly qualified (Tranel, 1994). Another exception is when a subpoena specifically squashes these terms of purchase, copyright, and trade secrets.

ASSESSING DIVERSE GROUPS

Competence in assessing diverse groups is an essential part of professional practice. This fact is highlighted by increased globalization, extensive immigration, controversies over potential test bias when used with diverse groups, cross-national adaptation of common instruments, and the American Psychological Association’s requirement that professional psychologists be trained to work with diverse groups. In the United States, as of 2013, more than one-third of the population was classified as members of an ethnic minority group (United States Census Bureau, 2015). Many minority populations in the United States are underrepresented and underserved (A. Levine, 2007). Thus, it is crucial that guidelines for competent assessment be developed. The guidelines pertain to language skills, cultural competency, assessing cultural/racial identity, appropriate use of instruments, diagnostic issues, and interpretation guidelines (see Dana, 2005).

Language Skills

Evaluating a client’s language proficiency is a first step in assessing diverse clients. Based on this evaluation, it may be necessary, or at least advisable, to conduct the assessment in the client’s native language. A sufficiently knowledgeable clinician can conduct the assessment him- or herself. Sometimes a translator or referral to another

clinician who speaks the language may be required. If the client is reasonably proficient in English, then it may be possible to conduct the assessment in English. However, clinicians should be aware of how this might alter the interaction and must take these potential differences into account when interpreting test scores. For example, a client who is struggling with English may appear to be uncooperative or to have flat affect when in reality this impression is created primarily because of language difficulties. It may also be advisable to use assessment instruments that have been translated into the client's native language.

Cultural Competency

Cultural competency on the part of clinicians begins with self-exploration of personal histories, attitudes, and knowledge. Doing this involves clinicians understanding their exposure to various cultures, biases about various cultures, and the degree of comfort with these cultures. It is natural to feel more resonant with some cultures as opposed to others. Often attitudes can be subtle and unconscious; for example, clinicians might have a sense of white privilege yet may have difficulty acknowledging these feelings. These attitudes are typically transmitted through nonverbal and subtle means.

Based on personal exploration and knowledge of a culture, clinicians need to develop optimal strategies of service etiquette. One strategy may involve level of formality. For example, Native Americans are likely to be more comfortable with minimal formality whereas Asian Americans usually expect more formal interactions characterized by a logical, structured approach. Other factors are the extent of eye contact, physical proximity, volume of voice, and the extent to which emotions are conveyed. For example, some cultures defer to persons perceived as being of higher status by decreasing the volume of their voice and minimizing eye contact. Clinicians who are knowledgeable about these differences should be both accepting of them and not misinterpret these behaviors as indicating depression or evasiveness. At the same time, these behaviors may make it more difficult to detect depression when it is actually present. A further variable is the time involved prior to the client becoming self-disclosing. Some cultures expect extensive preliminaries, perhaps to the point of having mutual acquaintances approve of the clinician prior to more formal clinical work. In contrast, other cultures are quite comfortable with becoming more self-disclosing and "task oriented" with minimal preliminaries. Taking into account each of these factors may make the difference between developing good rapport with accurate assessment results versus poor rapport resulting in inaccurate assessment data.

Cultural/Racial Identity

Cultural identity is a crucial aspect of explaining thoughts, feelings, and behaviors. It is thus important to understand this fact when conducting individual assessment. However, cultural identity varies according to the extent that a person identifies with his or her culture. Some individuals have quite strong identifications with their cultures. As a result, careful consideration of whether standard tests are appropriate to use with them is required. It may be necessary to use test translations, different norms, translators, or instruments specific to their culture. Further, cultural identity should be taken

into account when interpreting test scores. In contrast, other clients might have early experience with a culture but have later become quite acculturated into the dominant culture. As a result, standard tests might be used with more confidence.

Level of identity can be assessed informally through interview. There are also a variety of more formal instruments that ask questions related to variables such as language proficiency/preference, religious beliefs, foods, family structure, value orientation, socioeconomic status, collectivism/individualism, and culture-specific traditions, customs, and identifications. A sample of frequently used measures follows (see review by Dana, 2005):

African Americans: African American Acculturation Scale (Landrine & Klonoff, 1994)

Asian Americans: Asian Values Scale (Kim, Atkinson, & Yang, 1999)

Hispanic/Latinos: Acculturation Rating Scale for Mexican Americans (Cuellar, Arnold, & Maldonado, 1995)

Native Americans and Alaska Natives: Northern Plains Bicultural Immersion Scale (Allen, 1998)

One caution with these instruments is that sometimes individuals have quite different origins within the general group the instrument is trying to measure. This is particularly true for Hispanics and Asian Americans. For example, there are significant differences between Hispanics from Mexico and those from Argentina. Similarly, Japanese, Chinese, Koreans, and Hmong have many differences between their cultures. Despite this fact, a scale such as the Asian Values Scale is at least a start at looking at some of the common cultural values of these groups.

Test Equivalence and Appropriate Use of Instruments

Whether an instrument is culturally appropriate is based on a number of considerations, including the client's level of acculturation, language preference, language proficiency, availability of translations of the instrument, whether the construct is the same for the client's culture, availability of norms, and availability of possibly more appropriate alternatives specific to the client's culture. At the core of whether or not the test is appropriate is evaluating the equivalence of the test. Equivalence can be organized according to linguistic, conceptual, and metric equivalence. (See Table 2.1.)

If a test is not equivalent, it may result in bias against the group or individual it is evaluating. The term *bias in testing* refers to the presence of systematic error in the measurement of certain factors (e.g., academic potential, intelligence, psychopathology) among certain individuals or groups (Suzuki & Ponterotto, 2007). The possible presence of bias toward minority groups has resulted in one of the most controversial issues in psychological testing. More specifically, critics believe that psychological tests are heavily biased in favor of, and reflect the values of, European American, middle-class society. They argue that such tests cannot adequately assess intelligence or personality when applied to minority groups. Whereas the greatest controversy has arisen from the use of intelligence tests, the presence of cultural bias is also relevant in the use of personality testing. Over the past 20 years, discussion of bias has shifted from controversy

Table 2.1 Summary of Test Equivalence

Type	Definition	Issues/Strategies
Linguistic	Wording and content	Translate into new language and then retranslate again (“back-translate”), consider idioms and pictures, <i>adapt</i> not merely literal translation
Conceptual	Construct has same meaning	Same as construct validity, makes similar predictions, correlation, and factor analysis
Metric	Same psychometric features	Distributions, ranges, stability, comparable reliability, and validity, do scores mean the same things

over the nature and extent of bias to a more productive working through of how to make the most valid and equitable assessment based on current knowledge (see Dana, 2005; Geisinger, 2003; Handel & Ben-Porath, 2000).

The original controversies over test bias centered on determining whether tests are as valid for minority groups as for nonminorities. Differences often do exist in mean test scores; however, the meaning that can be attributed to these differences has been strongly debated. The major question lies in identifying the cause of these differences. Differences in test scores could stem from genuine differences in ability, which could be the result of environmental factors (Kamin, 1974; R. Rosenthal & Jacobson, 1968) or actual hereditary determination (A. R. Jensen, 1969, 1972; Rushton, 1994), or they could be artifacts of tests that are inherently biased. Although the debate is not resolved, guidelines have been established by the Equal Employment Opportunity Commission (EEOC, 1970) for the use of psychological tests with minority groups in educational and industrial settings. The basic premise is that a screening device (psychological test) can have an adverse impact if it screens out a proportionally larger number of minorities than nonminorities. Furthermore, it is the responsibility of the employer to demonstrate that the procedure produces valid inferences for the specific purposes for which the employer would like to use it. If an industrial or educational organization does not follow the guidelines as defined by the EEOC (1970), the Office of Federal Contract Compliance has the direct power to cancel any government contract that the institution might have.

Linguistic Equivalence

As summarized in Table 2.1, the first area of concern is *linguistic equivalence*, which is whether the test has been translated accurately. On the surface, this may mean simply translating the administration instructions and test items into the language of interest. One strategy to assist with this is to use “back-translation.” In back-translation, once the test is translated, it is then translated back into the original language. If the meanings of the items are still the same, then the back-translation helps to ensure that the translation is conceptually adequate. A further issue is that sometimes idioms need to be comparable. Similar to this issue is that not only verbal materials but also pictures should be made comparable. For example, a picture of a stereotypically appearing person depicted in one culture should be similarly made to look stereotypical in the culture

the test has been translated into. Doing this goes beyond merely translating the test and into an “adaptation” of the test (sometimes referred to as “functional equivalence”).

Conceptual Equivalence

A further concern is *conceptual equivalence*, which requires the constructs to have the same meaning in various cultures. Sometimes the equivalence of the constructs is clear, whereas other times it is more difficult to determine. For example, “dominance” as a personality trait may seem to be something that would be conceptually equivalent in all cultures. This fact is partially true, but nuances may make the concept somewhat different in various cultures. More collectivist cultures may emphasize the obligation to the group or family as being a more important aspect of dominance than individualistic cultures. It should be noted that various aspects of conceptual equivalence may emerge during translations of the test. For this reason, linguistic and conceptual equivalence are somewhat overlapping strategies.

More formal procedures for establishing conceptual equivalence might include investigating patterns of convergent and discriminant validity. A favored means of determining equivalence is factor analysis. It would be predicted that if indeed the concepts are comparable, then the same factors should emerge on the test when evaluated using samples from different cultures.

Metric Equivalence

The final means of establishing equivalence is through *metric equivalence*. This term refers to whether the instrument has similar psychometric properties across different groups/cultures. Assessing the extent to which the psychometric properties are different can include evaluating such areas as content, criterion, and construct validity. Note that a prerequisite for metric equivalence is that conceptual equivalence needs to be demonstrated first.

One of the initial things that persons reviewing tests will notice is that there are items on tests that appear irrelevant and possibly unfair for various groups. For example, a person from a different country could not reasonably be expected to know prominent political leaders in the country where the test was developed. On the surface, it would appear that such a test is culturally biased. Within the United States, early intuitive observations seemed to suggest that many African American children and other minorities usually do not have the opportunity to learn the types of material contained on many tests. Thus, their lower scores may represent not a lack of “intelligence,” but rather a lack of familiarity with European American, middle-class culture. Critics of tests point out that it would clearly be unfair to assess a European American’s “intelligence” based on whether he or she knows idioms or facts specific to a certain ethnic minority or national group. Low scores would simply measure an individual’s relative familiarity with the knowledge contained within the group rather than his or her specific “mental strengths.”

If this reasoning is used, many IQ and aptitude tests may appear on the surface to be culturally biased. However, studies in which researchers, to the best of their ability, eliminated biased test items or items that statistically discriminated between minorities and nonminorities did not alter total test scores (C. R. Reynolds, 2000).

In a representative study, 27 items were removed from the SAT that consistently differentiated minorities from nonminorities. This removal did little to change either the test takers' individual scores or the differences between the two groups (Flaughner & Schrader, 1978). Thus, the popular belief, based on a superficial appraisal of many psychological tests, that biased items are responsible for test differences does not appear to be supported by research.

Although test differences between minority and nonminority groups have frequently been found, the meaning and causes of these differences continue to be debated. For example, it has been demonstrated that African Americans consistently scored 12 to 15 IQ points lower than European Americans on the WISC-III and WAIS-III (Heaton, Taylor, & Manly, 2003; Prifitera, Weiss, & Saklofske, 1998). When African Americans and European Americans of equal socioeconomic status were compared, the differences in IQ scores were reduced to 11 to 13 IQ points (Heaton et al., 2003). Performance by Hispanics is about 7 IQ points lower than that of European Americans, and Asian Americans have been found to have IQ scores roughly equal to those of European Americans. Personality tests have also been found to have differences among various ethnic groups within the United States. For example, some studies (Dahlstrom, Lachar, & Dahlstrom, 1986; Timbrook & Graham, 1994) have found that African Americans have means five T-score points higher for MMPI scales F, 8, and 9. However, these differences were either decreased or found to be insignificant when groups were matched for age and education. This point suggests that socioeconomic factors may be an important reason for score differences. Socioeconomic status still accounts for only part of the reason for differences in test performance on cognitive tests, however (Sackett, Borneman, & Connelly, 2008). Other possible reasons are lack of belief in the impact of effort, level of acculturation, the effects of discrimination, gaps in general skills, or possible genetic differences. The reasons for these differences have been hotly debated and at this point are still unclear (see Neisser et al., 1996, and W. M. Williams, 2000).

Another consideration related to metric equivalence is the adequacy of the predictive validity of various tests when used with minority groups. Because one of the main purposes of these tests is to predict later performance, it is essential to evaluate the extent to which the scores in fact adequately predict areas such as performance in college for different populations. A representative group of studies indicates that SAT scores actually overpredict how well some minorities will perform in college (Hunter & Schmidt, 1996, 2000; A. R. Jensen, 1984; Sackett et al., 2008). Intelligence test scores have also been found to consistently predict African American work performance as accurately as European American performance (J. E. Hunter & Schmidt, 2000). Furthermore, both the WISC and the WISC-R were found to be equally as effective in predicting the academic achievement of both African Americans and European Americans in primary and secondary schools (Neisser et al., 1996; Reynolds & Hartlage, 1979).

A number of tests have been developed with the partial intent of using them in the assessment of ethnic minorities and cross-national groups, and they tend to emphasize nonverbal tasks. Included are the Leiter International Performance Scale, Peabody Picture Vocabulary Test-IV, Raven's Progressive Matrices, the Universal Nonverbal Intelligence Test-2, and the Test of Nonverbal Abilities (Bracken & McCallum, 2015;

McCallum, Bracken, & Wasserman, 2001). Some of these tests have been found to have minimal cultural bias (see Kaufman & Lichtenberger, 2006). In addition, the K-ABC-II (Kaufman et al., 2005) demonstrates minimal cultural bias. Mean IQ scores for European Americans, African Americans, and Hispanics are relatively close, and there is some evidence that reliability and concurrent validity is comparable for different ethnic populations (Kaufman et al., 2005).

As is true for ability tests and tests of scholastic aptitude, personality tests have the potential to be biased. The main research in this area has been performed on the MMPI/MMPI-2, and it has consistently indicated that minority groups do score differently than do nonminorities (see section titled "Use with Diverse Groups" in Chapter 7). However, these differences have not been found to be consistent across all populations (Greene, 1987, 1991, 2000). For example, African Americans from forensic, psychiatric, and vocational populations have been found to have varying patterns of mean scale elevations when compared to the mean scale elevations for European Americans. Even if consistent score differences were found, this does not mean these differences will be of sufficient magnitude to alter a clinician's interpretations, nor does it mean that predictions based on empirical criteria will be different. Studies using empirical criteria for prediction indicate that the MMPI does not result in greater descriptive accuracy for European Americans than for African Americans (Elion & Megargee, 1975; Greene, 1991, 2000).

Reviews of MMPI/MMPI-2 performance for Asian Americans, African Americans, Hispanics, and Native Americans have concluded that since no consistent patterns have emerged between various ethnic groups, it is premature to use different ethnically based norms (J. R. Graham, 2011; Greene, 1987, 1991; G. C. N. Hall, Bansal, & Lopez, 1999; Schinka, LaLone, & Greene, 1998). What seems to affect MMPI profiles more than ethnicity are moderator variables, such as socioeconomic status, intelligence, and education. Furthermore, the existing differences may result from true differences in behavior and personality caused by the greater stresses often encountered by minorities. J. R. Graham (1987) suggested that, when MMPI scores are deviant, the clinician should tentatively accept these scores but make special efforts to explore the person's life situation and level of adjustment and integrate this information with the test scores.

From this discussion, it should be obvious that developing test equivalence is complicated and that the meanings of various patterns of scores are far from being resolved. Several general solutions have been suggested (see Suzuki & Ponterotto, 2007). These include improving selection devices, developing different evaluation criteria, improving general skills, and changing social environments. Improving the use of selection devices involves paying continual attention to, and obtaining greater knowledge of, the meaning of different scores for different subgroups. Doing this may include tailoring specific test scores to the types of decisions individuals may make in their lives.

Another approach to solving the problem of potential test equivalence and bias is to develop different and more adequate criterion measures. For example, objective measures of work performance may be more accurate predictors than formal tests. These predictions of work performance may be higher if made by persons who share similar ethnic backgrounds. Related to this point, it may be crucial to consider the impact of various settings. For example, if a European American and a Latino attorney are placed in settings in which they work with Latinos, it is probable that the Latino

attorney would be more effective because he or she will have increased rapport and greater familiarity with the language and values of his or her clientele.

Another solution involves changing the social environment. Part of the rationale for emphasizing this approach is the belief that the differences in test scores between minorities and nonminorities are not because of test bias but rather because tests accurately reflect the effects of an unequal environment and unequal opportunities (C. R. Reynolds, 2000). Even though, in some situations, different minority norms and additional predictive studies on minority populations are necessary, the literature suggests that tests are not as biased as they have been accused of being (see Sackett et al., 2008). Removing seemingly biased or discriminating items still results in the same mean test scores, ability tests often still provide accurate predictions of grade point average for both minorities and nonminorities, and the MMPI-2 often makes behavioral predictions that are equally as accurate for various ethnic groups. These facts suggest that tests themselves are often not the problem but merely the means of establishing that, often, inequalities exist between ethnic groups. The goal should be to change unequal environments that can ideally increase a population's skills as measured by current tests of aptitude, IQ, and achievement. Whereas improving selection devices and developing different criterion measures are still important, future efforts should also stress more equal access to educational and other opportunities.

Probably the most important strategy is to maintain a flexible attitude combined with the use of alternative assessment strategies. Doing this changes the focus from merely establishing test equivalence to using a wide array of alternate assessment strategies. Thus, nonverbal techniques might be used, such as the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998), Raven's Progressive Matrices Test, or emphasis on the perceptual/nonverbal subtests of the WAIS-IV/WISC-V. In addition, "dynamic testing," in which actual observations of the benefit a person receives from learning situations, also shows promise in assessing the extent to which a client can benefit from educational interventions (learning potential; Grigorenko & Sternberg, 1998). Material beyond tests, such as teacher reports, discussions with parents, history, and behavioral observations, should also be given greater significance.

Diagnostic Issues

DSM-5 diagnosis needs to be considered within the context of cultural considerations. In addition to noting the cultural identity of the client, it is also crucial to carefully listen to cultural explanations of the client's difficulty. One category of presentation is the cultural concepts of distress that are outlined in the *DSM-5*. For example, *dhat syndrome* (mainly South Asia) is a cluster of symptoms that includes anxiety, depressive mood, and multiple somatic complaints. Also important is how the presence of oppression and discrimination among ethnic groups might contribute to misdiagnosing a person as being paranoid. A further example is how a disorder such as depression might be presented in primarily physiological terms within some cultures. In such cases, the external presentation would need to be decoded in order to identify the underlying depression. Research has clearly demonstrated varying rates of diagnoses in various cultures (Nguyen et al., 2007). What is less clear is whether these varying diagnoses represent genuine differences in rates or possible underdiagnosis, overdiagnosis,

or misdiagnosis. The practical implication is that when errors in diagnosis do occur, they have the potential to result in poor decisions and inappropriate treatment. It might also be necessary to consider combining standard psychological treatments with culture-specific interventions.

Interpretation Guidelines

The preceding discussion clearly indicates that ensuring accurate interpretations for diverse groups is challenging but also essential. Acculturation, equivalence, cultural competence, and the client's self-description within the context of his or her culture all need to be taken into account. Clinicians also need to incorporate what is known about how the instruments function within various cultures, including translations, idioms, norms, and various types of validity. However, it is nearly impossible to definitively demonstrate equivalence, due to the many steps and issues involved as well as the basic fact that error is inherent in any process aimed at equating two different cultures. Due to this fact, clinicians need to be both flexible and sensitive. For example, the pathological aspects of a high score on MMPI-2 Scale 6 (Paranoia) may need to be moderated if elevated for a client who has experienced significant racial discrimination. Similarly, indicators of low emotional expressiveness on the Rorschach may need to be modified if the person's emotional responses seemed to be "blunted" due to struggles with English as a second language. Often a phrase needs to be included in a report like "... results need to be treated with caution as the instruments have not been adequately adapted for use within the client's culture." Inserting a phrase such as this means that there are no clear specific strategies, but instead there are general guidelines to work with. Information and guidelines relevant to specific tests are included in each of the test-related chapters (see the sections titled "Use with Diverse Groups" in those chapters).

SELECTING PSYCHOLOGICAL TESTS

The most important factor in test selection is the extent to which the test is useful in answering the referral question. An assessment of neurological patients might use tests sensitive to cerebral deficit; patients with depression might be given the Beck Depression Inventory-II (A. T. Beck, Steer, & Brown, 1996); and patients with pain might be given the McGill Pain Questionnaire (Melzack, 1975), Millon Behavioral Health Inventory (Millon, Green, & Meagher, 2000), or Illness Behavior Questionnaire (Pilowski, Spence, Cobb, & Katsikitis, 1984).

Another important factor in test selection is a particular practitioner's training, experience, personal preferences, and familiarity with relevant literature. For example, a clinician who has received training in the MMPI-2 might be concerned about its ability to assess personality disorders and may rather choose to use an instrument such as the MCMI-IV (Millon, Grossman, & Millon, 2015). Clinicians might also select an instrument because it has practical efficiency in terms of time and economy (Groth-Marnat, 1999). Thus, they may wish to use simple behavioral predictions made by the client rather than use more expensive, time-consuming, and, quite possibly, less accurate tests (Shrauger & Osberg, 1981). Computer-assisted instruments may also

help to lower the cost of assessment, primarily by reducing direct practitioner time and achieving greater speed for scoring and hypothesis generation. A final crucial factor is that the assessment instrument should have good psychometric properties (see Hunsley & Mash, 2008).

The most frequently used assessment techniques are discussed in Chapters 3 to 13. Contact details for the major psychological distributors, along with a partial listing of tests they carry, are listed in Appendix A. Additional information on tests and assessments can be found by contacting various organizations that focus on assessment, listed in Appendix B. Various combinations of tests typically constitute a core battery used by clinicians. However, it is often necessary to expand such a core battery depending on the specifics of the referral question. Table 2.2 provides a listing of the domains for assessment along with relevant tests.

Although some of the tests described in Table 2.2 are thoroughly described in specific chapters dedicated to them, others may be relatively unfamiliar, and practitioners should obtain additional information on them. Various sources are available for information about these and other tests. Such sources can provide important information for deciding whether to obtain the tests and incorporate them into a battery. Probably the most useful is the *Mental Measurements Yearbook*, which contains a collection of critical test reviews that include evaluations of the tests and an overview on the tests. The *Nineteenth Mental Measurements Yearbook* was published in 2014 (Carlson, Geisinger, & Jonson, 2014), but it may be necessary to consult previous editions as not all tests are reviewed again in each new edition. The reviews are available in book form as well as online (*Mental Measurement Database*; see www.buros.org). *Tests in Print VIII* (L. L. Murphy, Geisinger, Carlson, & Spies, 2011) is associated with the *Mental Measurements Yearbook* but, rather than focusing on evaluating tests, lists information on each test, such as title, population it was designed for, available subtests, updating, author(s), and publisher. A further listing, description, and evaluation of tests can be found in Maddox (2003), *Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business* (5th ed.), which provides descriptive information on more than 3,500 tests. Practitioners interested in obtaining information on rating scales and other measures used in clinical practice might consult *Measures for Clinical Practice and Research: A Sourcebook* (Fischer & Corcoran, 2007). In *A Guide to Assessments That Work*, Hunsley and Mash (2008) present tests according to types of disorders and provide descriptions of these tests along with ratings of their psychometric properties. Neuropsychological tests are reviewed in the preceding resources as well as in Lezak and colleagues' (2012) *Neuropsychological Assessment*; Strauss, Sherman, and Spreen's (2006) *Compendium of Neuropsychological Tests*; and specialty journals in neuropsychology, particularly *Neuropsychology Review*. A careful review of the information included in these references will often answer questions clinicians might have related to a test's psychometric properties, usefulness, appropriateness for different populations, details for purchasing, and strengths and limitations. Most of the questions listed in Table 1.1 (see Chapter 1) can be answered by consulting the preceding resources.

An important and current trend in research and practice on psychological assessment is the use of tests to generate a treatment plan (Harwood, Beutler, & Groth-Marnat, 2011; Groth-Marnat & Davis, 2014; Jongsma, Peterson, & Bruce, 2006; Maruish, 2004; Wright, 2010). Indeed, a basic objective of psychological

Table 2.2 Assessment Instruments Relevant for Specific Response Domains**Cognitive Functioning****General functioning**

Mental Status Examination
 Mini-Mental State Examination (MMSE)

Intellectual functioning

Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV)
 Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V)
 Stanford-Binet–Fifth Edition (SB5)
 Kaufman Assessment Battery for Children–Second Edition (KABC-2)
 Woodcock-Johnson Psychoeducational Battery–Fourth Edition (WJ-IV)

Memory functioning

Wechsler Memory Scale–Fourth Edition (WMS-IV)
 Rey Auditory Verbal Learning Test
 California Verbal Learning Test (CVLT)
 Benton Visual Retention Test

Visuoconstructive abilities

Bender Visual Motor Gestalt Test–Second Edition (Bender-2)
 Drawing tests

Content of thought processes

Thematic Apperception Test (TAT)
 Children’s Apperception Test (CAT)
 Roberts’ Apperception Test for Children (RATC)

Academic Achievement

Woodcock Johnson Tests of Achievement–Fourth Edition (WJ-IV)
 Wide Range Achievement Test–Third Edition (WRAT-III)
 Wechsler Individual Achievement Test–Third Edition (WIAT-III)

Personality Functioning, Emotional Functioning, and Level of Psychopathology**General patterns and severity**

Minnesota Multiphasic Personality Inventory–Second Edition (MMPI-2)
 Minnesota Multiphasic Personality Inventory–Second Edition RF (MMPI-2-RF)
 Millon Clinical Multiaxial Inventory–Fourth Edition (MCMI-IV)
 Millon Adolescent Clinical Inventory (MACI)
 Rorschach
 Symptom Checklist 90–Revised (SCL-90)
 Brief Symptom Inventory (BSI)
 Personality Inventory for Children–Second Edition (PIC-2)

(continued)

Table 2.2 (Continued)

General personality measures
Sixteen Personality Factors (16-PF)
NEO-PI-R
Adjective Checklist
Sentence completion tests
Diagnosis
Diagnostic Interview Schedule
Schedule for Affective Disorders and Schizophrenia
Structured Clinical Interview for DSM (SCID)
Structured Interview for DSM Personality Disorders (SCID-2)
Diagnostic Interview for Children and Adolescents
Depression
Beck Depression Inventory–Second Edition (BDI-2)
Hamilton Rating Scale for Depression
Children’s Depression Inventory
Anxiety
State-Trait Anxiety Inventory
Fear Survey Schedule
Anxiety Disorders Interview Schedule
Sexual disturbance
Derogatis Sexual Functioning Inventory
Alcohol abuse
Michigan Alcoholism Screening Test
Alcohol Use Inventory
Interpersonal patterns
California Psychological Inventory (CPI)
Rathus Assertiveness Schedule
Therapeutic Reactance Scale
Taylor Johnson Temperament Analysis
Marital/family disturbance
Dyadic Adjustment Scale
Family Environment Scale
Marital Satisfaction Inventory
Academic/school adjustment
Achenbach Child Behavior Checklist (CBCL)
Vineland Social Maturity Scale
Connors Behavior Rating Scales
Kinetic School Drawing
Behavior Assessment System for Children–Second Edition (BASC-2)

Table 2.2 (Continued)

Adaptive level
AAMD Adaptive Behavior Scale
Vineland Adaptive Behavior Scale
Prognosis and risk
Suicide potential
Scale of Suicide Ideation
Beck Hopelessness Scale
Schizophrenia prognosis
Camberwell Family Interview
Vocational Interests
Career Assessment Inventory
Kuder Occupational Interest Survey
Self-Directed Search
Strong Interest Inventory (SII)

assessment is to provide useful information regarding the planning, implementation, and evaluation of treatment. With the increased specificity of both treatment and assessment, this goal is becoming possible. For example, oppositional, resistant clients have been found to have optimal treatment outcomes when either self-directed or paradoxical interventions have been used (Beutler, Clarkin, & Bongar, 2000; Beutler, Sandowicz, Fisher, & Albanese, 1996). In addition, a problem's severity has clear implications for the restrictiveness of treatment (inpatient, day treatment, outpatient) as well as treatment duration and intensity. Thus, clinicians should not select tests based simply on their diagnostic accuracy or psychometric properties; they should also be concerned with the functional utility of the tests in treatment planning. Accordingly, Chapter 14 presents a systematic, integrated approach to transforming assessment results into a series of clear treatment recommendations.

One special concern in selecting tests is faking. In many situations, clinicians might be concerned that persons will either consciously or unconsciously provide inaccurate responses (see kspope.com/assess/). Malingering ("inconsistent effort") is becoming an increasingly important issue, especially in forensic settings, where personal gain may result in presenting "fake bad" results. Thus, clinicians may want to pay particular attention to validity scales built in to tests (e.g., MMPI-2, MCMI-IV) and use specialty instruments designed to detect faking (e.g., Test of Memory Malingering, Structured Interview of Reported Symptoms). Although controversial, many projective techniques may be resistant to attempts at faking.

Another special concern in selecting tests relates to the time required for assessment, which may cause examiners to consider selecting short forms of instruments such as the WAIS-IV or WISC-V. Although many short forms for cognitive tests seem sufficiently valid for screening purposes, their use as substitutes for the longer forms is not acceptable (Kaufman, Kaufman, Balgopal, & McLean, 1996; Kaufman & Lichtenberger, 2002). Most past attempts to develop short forms for the longer

objective personality tests, such as the MMPI-2, have not been successful (Butcher, 2011). However, future computerized applications that tailor items based on a client's previous responses (adaptive testing) may result in the development of shortened administrations with acceptable psychometric properties (Forbey & Ben-Porath, 2007). In addition, the recent 338-item MMPI-2 Restructured Form is a shorter and psychometrically improved version (Ben-Porath & Tellegen, 2008/2011).

During the evaluation of single cases, such as in clinical diagnosis and counseling, clinicians do not usually use formal combinations of test scores. Rather, they rely on their past judgment, clinical experience, and theoretical background to interpret and integrate test scores. However, for personnel decisions, academic predictions, and some clinical decisions (recidivism rate, suicide risk), clinicians may be advised to use statistical formulas (Aegisdottir et al., 2006). The two basic approaches for combining test results are multiple regression equations and multiple cutoff scores. Multiple regression equations are developed by correlating each test or subtest with a criterion. The higher the correlation, the greater is the weight in the equation. The correlation of the entire battery with the criterion measure gives an indication of the battery's highest predictive validity. For example, high school achievement can be predicted with this regression equation, which combines IQ and California Psychological Inventory (CPI) subtests:

$$\begin{aligned} \text{Achievement} = & .786 + .195 \text{ Responsibility} + .44 \text{ Socialization} \\ & - .130 \text{ Good Impression} + .19 \text{ Achievement via Conformance} \\ & + .179 \text{ Achievement via Independence} + .279 \text{ IQ} \end{aligned}$$

This equation raises the correlation with grade point average to .68 as compared with .60 when using IQ alone (Megargee, 1972). This correlation indicates that academic achievement is dependent not only on intellectual factors but also on psychosocial ones, such as responsibility, socialization, achievement via independence, and achievement via conformance, all of which are measured by the CPI. The second strategy, multiple cutoff scores, involves developing an optimum cutoff for each test or subtest. If the person is above a certain specified score (e.g., above the range for brain damage or schizophrenia), the score can be used to indicate the presence of a certain characteristic. Although equations or cutoffs have not been developed for all tests, the decision to include a test in a battery may depend in part on the presence of such formal extensions of the test. In addition, many of the computer-assisted interpretive packages use various actuarial formulas (usually in combination with expert interpretations) to develop their interpretations.

COMPUTER-ASSISTED ASSESSMENT

During the past 40 years, computer-assisted assessment has grown exponentially. By 1990, 17% of practicing psychologists frequently used computer-generated narratives, with an additional 36% using them on an occasional basis (Spielberger & Piotrowski, 1990). By 1999, the number of psychologists stating that they used some form of computer-assisted testing had increased to 40% (McMinn, Buchanan et al., 1999). More than 400 software packages are available, many of which are listed in various

catalogs published and distributed by test suppliers. At present, computers are used mainly for their clerical efficiency in scoring and data storage and to generate interpretive reports. However, more and more, testing is becoming available in computer/tablet-assisted formats. Future uses of computers are likely to continue to experiment with features such as innovative presentation of items (e.g., adaptive testing), networked norms, novel presentation of stimuli (e.g., virtual reality), psychophysiological monitoring, and artificial intelligence (Garb, 2000; Groth-Marnat, 2000a, 2009). Computing in mental health has included not only computer-assisted assessment but also computer interviews, computerized diagnosis, computer-aided instruction, direct treatment intervention, clinical consultation, and simulated psychiatric interviews (Lichtenberger, 2006; McMin, Buchanan et al., 1999).

There have been a number of particular advances in computer-assisted administration and interpretation in neuropsychology (see special series review by Kane, 2007). Batteries have been developed mainly in large organizational contexts (military, Federal Aviation Authority) and focused on specialized types of problems. For example, the Neurobehavioral Evaluation System (NES) is particularly sensitive to the impact of environmental toxins (Groth-Marnat, 1993), CogScreen has been used in the selection of airline pilots, and the military's Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB) was originally developed to assess the impact of drugs in the workplace. The Cambridge Neuropsychological Test Automated Batteries (CANTAB) have been found to detect and locate brain damage including early signs of Alzheimer's, Parkinson's, and Huntington's diseases (Fray, Robbins, & Sahakian, 1996; Luciana, 2003). Although computer-assisted programs show considerable promise, they are currently used less than the more familiar individually administered neuropsychological tests or test batteries (Camara, Nathan, & Puente, 2000; Luciana, 2003).

Computer-assisted assessment has a number of advantages. Use of computers can save valuable professional time, potentially improve reliability and fidelity to standardized administration, reduce possible tester bias, and reduce the cost to the consumer by improving efficiency (Butcher, Perry, & Hahn, 2004; Groth-Marnat, 1999; Kane, 2007; Luciana, 2003). Even greater benefits may someday be realized by incorporating more complicated decision rules in interpretation, collecting data on response latency and key pressure, incorporating computer-based models of personality, tailoring future questions to a client based on past responses, and estimating the degree of certainty of various interpretations (Groth-Marnat, 2000a, 2000b; Lichtenberger, 2006).

In the past, computer-assisted assessment has resulted in considerable controversy within mental health publications (Faust & Ziskin, 1989; Groth-Marnat & Schumaker, 1989), the popular media (C. Hall, 1983), and professional publications outside the mental health area (Groth-Marnat, 1985). A primary issue has been untested reliability and validity. Research on reliability, however, has typically indicated that computerized administrations have generally excellent reliability that is at least equivalent to the paper-and-pencil versions (Campbell et al., 1999; Kane, 2007; Luciana, 2003). In addition, computer-administered versus paper-and-pencil outcomes for traditional tests have generally been found to result in negligible differences in scores (Butcher et al., 2004; Finger & Ones, 1999). This finding supports the view that if a paper-and-pencil version of the test is valid, a computerized version will have equal validity resulting from the comparability in scores.

A further issue is the validity of computer-based test interpretation. Butcher et al. (2004) concluded that in the vast majority of computer-based interpretations, 60% of the interpretations were appropriate. Shorter to mid-length narratives were generally considered to have a higher proportion of valid interpretations when compared with longer ones. In addition, the narrative statements contained in the computer-based reports were comparable to the types of statements made by clinicians. Although this finding generally supports the use of computer-based interpretations, the fact that 40% or more of interpretations were not considered accurate means that the computer-based interpretations should be carefully evaluated. Thus, cutting and pasting computerized narratives into reports results in unacceptably high error rates. Indeed, 42% of psychologists surveyed felt this procedure raised ethical concerns (McMinn, Ellens et al., 1999). The previous summary clearly emphasizes that computer-based reports should not be used to replace clinical judgment but should instead be used as an adjunct to provide possible interpretations for the clinician to consider.

The Association of Test Publishers (2000) attempted to clarify standards in its *Guidelines for Computer-Based Testing* (as did the 2010 American Psychological Association's ethics code). The association stressed that only persons who meet the requirements for using psychological tests in general should use computer-based assessments (Turner, DeMers, Fox, & Reed, 2001). Specifically, users should have an understanding of psychological measurement, validation procedures, and test research. They should also limit their use of computerized techniques to those areas they are competent to use. They should be knowledgeable regarding how computer-based scores were generated and how interpretations have been made. Finally, they should be able to evaluate whether the computer-based procedures are applicable to how they will be used.

The preceding difficulties associated with computer-assisted assessment suggest a number of guidelines for users (Butcher et al., 2004; Groth-Marnat & Schumaker, 1989). First, practitioners should not blindly accept computer-based narrative statements but rather should ensure, to the best of their ability, that the statements are both linked to empirically based research and placed in the context of the unique history and unique situation of the client. Computers have, among other benefits, the strong advantage of offering a wide variety of possible interpretations to the clinician, but these interpretations still need to be critically evaluated. Far greater research needs to be performed on both the meaning of computer-administered test scores and on the narrative interpretations based on these scores. The developers of software should also be encouraged to provide enough information in the manual to allow proper evaluation of the programs and should develop mechanisms to ensure that obsolete programs are updated.

RECOMMENDED READING

- Butcher, J. N., Perry, J. N., & Hahn, J. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology*, 60, 331–345.
- Dana, R. H. (2005). *Multicultural assessment: Principles, applications, and examples*. Mahwah, NJ: Erlbaum.

- Heilbrun, K., Marczyk, G. G., & Dematteo, D. (2002). *Forensic mental health assessment: A casebook*. New York, NY: Oxford University Press.
- Hunsley, J., & Mash, E. J. (Eds.). (2008). *A guide to assessments that work*. New York, NY: Oxford University Press.
- Pope, K. (2007a). Informed consent in psychotherapy and counseling: Forms, standards & guidelines, & references. Retrieved from <http://kspope.com/consent/index.php>.
- Pope, K. (2007b). Responsibilities in providing psychological test feedback to clients. Retrieved from <http://kspope.com/assess/feedabs1.php>.
- Yalof, J., & Brabender, V. (2001). Ethical dilemmas in personality assessment courses: Using the classroom for in vivo training. *Journal of Personality Assessment*, 77, 203–213.
- Zuckerman, E. L. (2003). *The paper office: Forms, guidelines, and resources to make your practice work ethically, legally, and profitably*. New York, NY: Guilford Press.

THE ASSESSMENT INTERVIEW

The single most important means of data collection to provide context for psychological evaluation is the assessment interview. Without interview data, most psychological test results are meaningless. The interview also provides potentially valuable information that may be otherwise unobtainable, such as behavioral observations, idiosyncratic features of the client, and the person's reaction to his or her current life situation. In addition, interviews are the primary means for developing rapport.

Sometimes an interview is mistakenly thought to be simply a conversation. In fact, an interview and a conversation differ in many ways. An interview typically has a clear sequence and is organized around specific, relevant themes, because it is meant to achieve defined goals. Its general objectives are to gather information that cannot easily be obtained through other means, establish a relationship that is conducive to obtaining the information, develop greater understanding in both the interviewer and interviewee regarding problems, and provide direction and support in helping the interviewee deal with problems. The interviewer must have knowledge about the areas to be covered during the interview and direct and control the interaction to achieve specific goals.

A basic dimension of an interview is its degree of structure. Some interviews allow the participants to freely drift from one area to the next, whereas others are highly directive and goal oriented, often using structured ratings and checklists. The more unstructured formats offer flexibility, possibly higher rapport, the ability to assess how clients organize their responses, and the potential to explore unique details of a client's history. Unstructured interviews, however, have received frequent criticism, resulting in widespread distrust of their reliability and validity. As a result, highly structured and semistructured interviews have been developed that provide sound psychometric qualities, the potential for use in research, and the ability to be administered by less trained personnel.

Regardless of the degree of structure, any interview needs to accomplish specific goals, such as assessing the client's strengths, level of adjustment, the nature and history of the problem, diagnosis, and relevant personal and family history. Techniques for accomplishing these goals vary from one interviewer to the next. Most practitioners use at least some structured aids, such as intake forms that provide identifying data and basic elements of history. Obtaining information through direct questions on intake forms frees the clinician to investigate other aspects of the client in a more flexible, open-ended manner. Clinicians might also use a checklist to help ensure that they have covered all relevant areas. Other clinicians use one of the formally developed structured interviews, such as the Schedule for Affective Disorders and Schizophrenia (SADS) or Structured Clinical Interview for the DSM-IV (SCID).

HISTORY AND DEVELOPMENT

Early Developments

The earliest method of obtaining information from clients was through clinical interviewing. At first, these interviews were modeled after question-and-answer medical formats, but later, the influence of psychoanalytic theories resulted in a more open-ended, free-flowing style. Parallel to the appearance of the psychoanalytically oriented interview was the development of the more structured and goal-oriented mental status examination, originally formulated by Adolf Meyer in 1902. The mental status examination assessed relevant areas of a client's current functioning, such as general appearance, behavior, thought processes, thought content, memory, attention, speech, insight, and judgment. Professionals also expressed early interest in the relationship between biographical data and the prediction of occupational success or prognosis for specific disorders.

Regardless of the style used, the interviews all had these common objectives: to obtain a psychological portrait of the person, to conceptualize what is causing the person's current difficulties, to make a diagnosis, and to formulate a treatment plan. The difficulty with unstructured interviews is that they were (and still are) considered to have questionable reliability, validity, and cost-effectiveness. The first standardized psychological tests were developed to overcome these limitations. Tests could be subjected to rigorous psychometric evaluation and were more economical because they required less face-to-face contact with the person(s) being evaluated.

Developments during the 1940s and 1950s

During the 1940s and 1950s, researchers and clinicians began conceptualizing and investigating five critical dimensions of interviews:

1. Content versus process
2. Goal orientation (problem solving) versus expressive elements
3. Degree of directiveness
4. Amount of structure
5. The relative amount of activity expressed by the participants

These issues have been the focus of numerous research studies. A representative and frequently cited study on interviewer style was reported by W. Snyder (1945), who found that a nondirective approach was most likely to create favorable changes and self-exploration in clients. In contrast, a directive style using persuasion, interpretation, and interviewer judgment typically resulted in clients being defensive and resistant to expressing difficulties. Strupp (1958) investigated the experience-inexperience dimension and found, among other things, that experienced interviewers expressed more warmth, a greater level of activity, and a greater number of interpretations. Level of empathy did not differ based on the interviewer's degree of experience. Further representative studies include Porter's (1950) in-depth evaluation of the effects of different types of responses (evaluative, probing, reassuring) and R. Wagner's (1949) early review, which questioned the reliability and validity of employment interviews.

Developments During the 1960s

A considerable amount of research in the 1960s was stimulated by C. Rogers (1961), who emphasized understanding the proper interpersonal ingredients necessary for an optimal therapeutic relationship (warmth, positive regard, genuineness). Elaborating on Rogers's ideas, Truax and Carkhuff (1967) developed a 5-point scale to measure interviewer understanding of the client. This scale was used for research on interviewing and therapist training and as support for a client-centered theoretical orientation. Additional research efforts were directed toward listing and elaborating on different categories of interactions, such as clarification, summarizing, and confrontation.

Other investigators conceptualized interviewing as an interactive system in which the participants simultaneously influenced each other (Matarazzo, 1965; Watzlawick, Beavin, & Jackson, 1966). This emphasis on an interactive, self-maintaining system became the core for most early and later formulations of family therapy. The 1960s also saw the development and formalization of behavioral assessment, primarily in the form of goal-directed interviews that focused on understanding current and past reinforcers, as well as on establishing workable target behaviors. Proponents of behavioral assessment also developed formal rating instruments and self-reports for areas such as depression, assertiveness, and fear.

Some attempts were made at integrating different schools of thought into a coherent picture, such as Beier's (1966) conceptualization of unconscious processes being expressed through nonverbal behaviors that could then be subject to covert social reinforcement. However, the 1960s (and part of the 1970s) were mostly characterized by a splintering into different schools of conflicting and competing ideologies. For example, client-centered approaches emphasized the importance of staying with the client's self-exploration; behavioral interviews emphasized antecedents and consequences of behavior; and family therapy focused on interactive system processes. Parallel progress was made within each of these different schools and within different disciplines, but little effort was devoted to cross-fertilization and/or integration.

Throughout the 1950s and 1960s, child assessment was conducted primarily through interviews with parents. Direct interviews with the child were considered to be for therapeutic purposes rather than for assessment. Differential diagnosis was unusual; almost all children referred to psychiatric clinics were either undiagnosed or diagnosed as "adjustment reactions" (Rosen, Bahn, & Kramer, 1964). Early research by Lapouse and Monk (1958, 1964) using structured interviews indicated that mothers were more likely to report overt behaviors that are bothersome to adults (thumb-sucking, temper tantrums), but children were more likely to reveal covert difficulties (fears, nightmares). Somewhat later, P. Graham and Rutter (1968), using structured interviews of children (rather than a parent), found interrater agreement was high for global psychiatric impairment (.84); moderate for attentional deficit, motor behavior, and social relations (.61–.64); and low for more covert difficulties, such as depression, fears, and anxiety (.30).

Developments During the 1970s

Assessment with adults and children during the 1970s saw a further elaboration and development of the trends of the 1960s, as well as increased emphasis on structured

interviews. The interest in structured interviews was fueled largely by criticisms about the poor reliability of psychiatric diagnosis. Typical structured interview data would be transformed into such scales as organicity, disorganization, or depression-anxiety.

Initial success with adult structured interviews (e.g., Present State Examination, Renard Diagnostic Interview) encouraged thinking regarding the further development of child-structured interviews both for global ratings and for specific content areas. Child assessment became concerned not only with information derived from parents but also with the child's own experience. There was a trend toward direct questioning of the child, greater emphasis on differential diagnosis, and the development of parallel versions of structured interviews for both the parent(s) and child.

Behavioral strategies of interviewing for both children and adults not only emphasized the interviewee's unique situation but also provided a general listing of relevant areas for consideration. Kanfer and Grimm (1977) outlined the areas an interviewer should assess as:

1. Behavioral deficiencies,
2. Behavioral excesses,
3. Inappropriate environmental stimulus control,
4. Inappropriate self-generated stimulus, and
5. Problem reinforcement contingencies.

In a similar categorization, Lazarus (1973, 2005) developed his BASIC-ID model, which described a complete assessment as involving *behaviors, affect, sensation, imagery, cognition, interpersonal relations*, and need for pharmacological intervention/drugs.

Additional themes in the 1970s included interest in biographical data, online computer technology, and the training of interviewer skills. Specifically, efforts were made to integrate biographical data for predicting future behavior (suicide, dangerousness, prognosis for schizophrenia) and for inferring current traits. J. W. Johnson and Williams (1980) were instrumental in developing some of the earliest online computer technology to collect biographical data and to integrate it with test results. Although training programs included interviewing skills, a central debate was whether these skills could actually be significantly learned or improved (Wiens, 1976).

Whereas most reviews of the literature in the 1970s emphasized the advantages of a comprehensive structured format, family therapists were dealing with group processes in which formal interview structure was typically deemphasized. Because most family therapists were observing fluid interactional processes, they needed to develop a vocabulary different from that used in traditional psychiatric diagnosis. In fact, *DSM* categories were usually considered irrelevant because they described static characteristics of individuals rather than ongoing group processes. Few, if any, structured formats were available to assess family relationships.

Developments During the 1980s

Many of the trends, concepts, and instruments developed in the 1960s and 1970s were further refined and adapted for the 1980s. One important effort was the adaptation

of many instruments to the *DSM-III* (1980) and *DSM-III-R* (1987). In addition, the increased delineation of childhood disorders required greater knowledge related to differential diagnosis and greater demand for structured interviews as adjuncts to assessment. Many of the efforts were consistent with the use of specific diagnostic criteria, along with a demand for efficiency, cost-effectiveness, and accountability. Despite concerns regarding computer-based interpretations (Groth-Marnat & Schumaker, 1989), some of these functions were beginning to be performed by specific computer programs. Because interviews were becoming increasingly structured, with the inclusion of scales and specific diagnostic strategies, the distinction between tests and interviews was becoming less clear. In some contexts, aspects of interviewing were even replaced with computer-requested and computer-integrated information and combined with simple programs to aid in diagnosis, such as DIANO III (Spitzer, Endicott, & Cohen, 1974) and CATEGO (Wing, Cooper, & Sartorius, 1974). During the mid- and late 1980s, most clinicians, particularly those working in large institutions, used a combination of structured interviews and open-ended unstructured approaches. Some research focused on the importance of the initial interview regarding clinical decision making and later therapeutic outcome (Hoge, Andrews, Robinson, & Hollett, 1988; Turk & Salovey, 1985). There was also a greater appreciation and integration of the work from different disciplines and from differing theoretical persuasions (Bellack & Hersen, 1988). Finally, greater emphasis was placed on the impact and implications of culture and gender on the assessment process (L. Brown, 1990).

The 1990s and Into the Millennium

Two of the defining features of psychology in the 1990s were managed health care and the controversy over the validity of repressed memories. Both of these issues had significant implications for interviewing. Managed health care emphasized the cost-effectiveness of providing health services; and for interviewing, this means developing the required information in the least amount of time. Doing this may mean streamlining interviews by maximizing computer-derived information or self-administered forms. The use of computer-assisted interviewing brings up the larger issue of the extent to which practitioners need to spend face-to-face time with the client rather than deriving information through other means. The development of single-session therapy (Hoyt, 1994) illustrates the potential brevity of information gathering that might be required before making therapeutic interventions. There was also recognition that precise patient–treatment matching can optimize the treatment and potentially the cost-effectiveness of psychosocial interventions (Antony & Barlow, 2011; Beutler & Clarkin, 1990; Beutler, Clarkin, & Bongar, 2000).

The controversy over repressed memories has forced interviewers to clarify the extent to which the information they derive from clients represents literal as opposed to narrative truth. Research has consistently indicated that client self-reports are reconstructions of events (Henry, Moffitt, Caspi, Langley, & Silva, 1994; Loftus, 1993) and are likely to be particularly questionable for retrospective reports of psychosocial variables (Garb, 2007; Henry et al., 1994; Piasecki, Hufford, Solhan, & Trull, 2007). The even greater challenge to interviewers is to ensure that their interviewing style and method of questioning are not distorting the information derived from clients.

This issue becomes intensely highlighted during interviews to investigate the possibility of childhood sexual abuse (see guidelines in S. White & Edelstein, 1991).

Further themes in the 1990s and into the millennium were the importance of interview strategies for special populations and the development of new technologies. It is clear that many diverse populations are more likely to be misdiagnosed. At least in part, this misdiagnosis results in worse outcomes compared with majority groups (Neighbors et al., 2007; Nguyen, Huang, Arganza, & Liao, 2007). The potential for misdiagnosis for minority groups demands that clinicians be aware of their own biases, become knowledgeable regarding these subgroups, and make appropriate modifications to their interviews (Ponterotto & Grieger, 2007). Several new technologies are both available and becoming progressively more utilized. These include computer-administered interviews (Garb, 2007) as well as data derived from electronic diaries (Piasecki et al., 2007) and ambulatory sensors (Haynes & Yoshioka, 2007) that become a part of clinical interviews. The themes and issues related to cost-effectiveness, patient–treatment matching, recovered memories, use of new interview technologies, and strategies for interviewing special populations will continue to be important themes throughout the first few decades of the millennium.

ISSUES RELATED TO RELIABILITY AND VALIDITY

Although the interview is not a standardized test, it is a means of collecting data and, as such, can and should be subjected to some of the same types of psychometric considerations as formal tests. Evaluating the psychometric properties of interviews is important because interviews can introduce numerous sources of bias, particularly if the interviews are relatively unstructured. Reliability of interviewers is usually discussed in relation to interrater (interviewer) agreement. R. Wagner's (1949) early review of the literature found tremendous variation, ranging from .23 to .97 (*Mdn*.57) for ratings of personal traits and .20 to .85 (*Mdn*.53) for ratings of overall ability. Later reviews have generally found similar variations in interrater agreement (Arvey & Campion, 1982; L. Ulrich & Trumbo, 1965). The problem then becomes how to determine which ratings to trust and which to view with skepticism. Of particular relevance is why some interviewers focus on different areas and have different biases. A consistent finding is that, when interviewers were given narrow areas to assess and were trained in interviewer strategies, interrater agreement increased (Dougherty, Ebert, & Callender, 1986; Zedeck, Tziner, & Middlestadt, 1983). The consensus is that highly structured interviews were more reliable (Garb, 2007; Huffcutt & Arthur, 1994). However, increased structure undermines one of the greatest strengths of interviews—their flexibility. In many situations, a free-form, open-ended approach may be the best way to obtain some types of information.

Research on interview validity has typically focused on various sources of interviewer bias. Halo effects result from the tendency of an interviewer to develop a general impression of a person and then infer other seemingly related characteristics. For example, clients who are considered to express warmth may be seen as more competent or mentally healthy than they actually are. This clustering of characteristics may be incorrect, thereby producing distortions and exaggerations. Similarly, first

impressions have been found to bias later judgments (W. Cooper, 1981). Confirmatory bias might occur when an interviewer makes an inference about a client and then directs the interview to elicit information that confirms the original inference. This bias typically occurs when clinicians develop initial diagnostic impressions and then ignore later relevant information since they are somehow invested in confirming their initial impressions. Similarly, a psychoanalytically oriented interviewer might focus on questions related to early childhood traumas, possibly incorrectly confirming traditional psychoanalytic explanations of current adult behaviors. Similar to halo effects is the finding that one specific outstanding characteristic (e.g., educational level, physical appearance) can lead an interviewer to judge other characteristics that he or she incorrectly believes are related to the outstanding one. For example, physical attractiveness has been found to create interviewer bias in job applicants (Gilmore, Beehr, & Love, 1986). In a clinical context, physical attractiveness may result in practitioners either deemphasizing pathology or, on occasion, exaggerating pathology because of discomfort the interviewers may feel over their feelings of attraction (L. Brown, 1990). Interviewers also may focus incorrectly on explanations of behavior that emphasize traits rather than situational determinants (Ross, 1977).

In addition to the interviewer's perceptual and interactional biases, interviewees themselves may distort their responses. Some specific areas of distortions include victims of automobile accidents typically exaggerating the amount of time they lost from work; 40% of respondents providing overestimates of their contributions to charity; and 17% of respondents reporting their ages incorrectly (R. Kahn & Cannell, 1961). Some interviewees may present an overly favorable view of themselves, even if they are relatively naive regarding their motivations. Distortions, however subtle, are often found in sensitive areas, such as sexual behavior. More extreme cases of falsification occur with outright (conscious) lies, delusions, confabulations, and lies by pathological (compulsive) liars that they partially believe themselves (Kerns, 1986). Inaccuracies based on retrospective accounts have been found to most likely occur related to psychosocial information (e.g., family conflict, onset of psychiatric symptoms) compared with variables such as change of residence, reading skill, height, and weight (B. Henry et al., 1994).

Reviews of interview validity, in which interviewer ratings were compared with outside criterion measures, have, like reliability measures, shown tremendous variability, ranging from $-.05$ to $+.75$ (Arvey & Campion, 1982; Henry et al., 1994; Huffcutt & Arthur, 1994; J. Hunter & Hunter, 1984; L. Ulrich & Trumbo, 1965). One clear finding is that validity increases as the structure of the interview format increases (Huffcutt & Arthur, 1994; Marchese & Muchinsky, 1993). For example, a meta-analysis by Wiesner and Cronshaw (1988) found that unstructured interviews had validity coefficients of $.20$, structuring the interview increased the validity to $.63$, and structured interviews by a panel using consensus ratings increased validity coefficients to a quite respectable $.64$. However, the validity seems to vary according to the type of variable that is being assessed. Situational employment interviews (i.e., asking the interviewee what he or she would do in a particular situation) had higher validities ($.50$) than interviews used to assess past job-related behavior ($.39$) or rate psychological qualities such as dependability ($.29$; McDaniel, Whetzel, Schmidt, & Maurer, 1994). It has also been found that interview accuracy increases more when interviewees are held accountable for the

process they went through when coming to their decisions, compared to being held accountable for the accuracy of their predictions (procedural versus outcome accountability; Brtek & Motowidlo, 2002).

The previous brief review indicates that adding structure to interviews and paying close attention to the procedure by which decisions are made typically result in higher levels of validity. It also means that information derived from unstructured interviews should be treated cautiously and treated simply as hypotheses that need to be supported by other means. Interviewers should also continually question the extent to which their particular style, attitudes, and expectations might be compromising interview validity. Given the difficulties related to unstructured formats, a variety of formal structured clinical interviews have been developed. Additional information on the reliability and validity of the most frequently used structured clinical interviews is provided in the “Structured Interviews” section of this chapter.

ASSETS AND LIMITATIONS

Both structured and unstructured interviews allow clinicians to place test results in a wider, more meaningful context. In addition, biographical information from interviews can be used to help predict future behaviors; what a person has done in the past is an excellent guide to what he or she may continue doing in the future. Improving prediction of suicide risk, success in certain occupations, and prognosis for certain disorders can often be effectively accomplished by attending to biographical data rather than test scores.

Because tests are almost always structured or “closed” situations, the unstructured or semistructured interview is typically the only time during the assessment process when the clinician can observe the client in an open, ambiguous situation. Observations can be made regarding how individuals organize their responses, and inferences can be derived from subtle, nonverbal cues. These inferences can be followed up with further, more detailed questioning. This flexibility inherent in unstructured and semistructured interviews is frequently their strongest advantage over standardized tests. The focus during unstructured interviews is almost exclusively on the individual rather than on how that individual does or does not compare with a larger normative comparison group. Some types of information can be obtained only through this flexible, person-centered approach, which allows the interviewer to pay attention to idiosyncratic factors. In crisis situations when relatively rapid decisions need to be made, it can be impractical to take the time required to administer and interpret tests, leaving interviews and rapid screening devices as the only means of assessment. Finally, interviews allow clinicians to establish rapport and encourage client self-exploration. Rarely do clients reveal themselves or perform optimally on tests unless they first sense trust, openness, and a feeling of being understood.

The greatest difficulty with unstructured interviews is interviewer bias from perceptual and interactional processes such as the halo effect, confirmatory bias, and the primacy effect. This bias typically results in considerable variability for both reliability

and validity, as well as in difficulty comparing one subject with the next. One of the main reasons for diagnostic disagreement is variations in the information obtained (information variance) and variations in the criteria (criterion variance) used to conclude the presence or absence of a condition. Variation in interviewing means that different practitioners develop and ask a wide variety of questions and apply standards for the presence of a condition, such as depression, in an inconsistent fashion.

Structured interviews have many distinct advantages over unstructured approaches. Because structured interviews have more psychometric precision, the results enable comparability between one case and the next (or the population). The standardized presentation allows for the development of reliable ratings, reduces information variance, and uses consistent diagnostic criteria (Garb, 2007; Summerfeldt & Antony, 2002). In addition, the comprehensiveness of many structured interviews reduces the likelihood of missing a diagnosis or set of relevant symptoms. Partly because of these advantages, structured clinical interviews have progressed from being used primarily for research to use in a number of clinical settings. One issue, however, is the time required for structured interviews. The more recently developed, but not widely used, computer-assisted programs offer a potential method of countering this difficulty (Epstein & Klinkenberg, 2001; Garb, 2007). In addition, computer-administered interviews are comprehensive, and clients are more likely to disclose highly sensitive information when compared with clinician-administered interviews (Garb, 2007). Instruments such as the Diagnostic Interview Schedule (DIS) and Diagnostic Interview for Children and Adolescents (DICA) have been designed for administration by lay interviewers, thereby reducing the time required by professionals.

Although structured interviews generally have stronger psychometric properties than unstructured formats, they tend to overlook the idiosyncrasies and richness of the person. In many cases, these unique aspects may go undetected and yet may make a significant difference in interpreting test scores or making treatment recommendations. Although still somewhat controversial (Helzer & Robins, 1988), another criticism by many clinicians and researchers is that highly structured approaches may not create enough rapport for the client to feel sufficiently comfortable about revealing highly personal information. This is truer for the highly structured interviews, such as the DIS, than for a semistructured instrument, such as the SADS, which includes an initial, relatively unstructured component. However, M. Rosenthal (1989) noted that rapport with structured instruments can be enhanced through carefully educating the client as to the importance and procedures of these more structured approaches.

Although many of the structured interviews have demonstrated adequate reliability, studies relating to validity have primarily focused on the general level of impairment or simple discriminations between psychiatric and nonpsychiatric populations. There has been considerable controversy over what exactly is an acceptable outside criterion measure regarding the “true” diagnosis. In-depth studies of construct validity or incremental validity have yet to be performed. Furthermore, far more work needs to be done on the treatment utility of structured interviews in areas such as selection of treatment, likely response to specific forms of pharmacological or psychotherapeutic interventions, and prognosis.

THE ASSESSMENT INTERVIEW AND CASE HISTORY

General Considerations

The previously mentioned historical and psychometric considerations indicate that no single correct way exists to conduct an unstructured or semistructured interview. Interviewer style is strongly influenced by theoretical orientation and by practical considerations. Persons strongly influenced by client-centered theories tend to be nondirective and to avoid highly structured questions. This is consistent with the underlying belief that persons have the inner ability to change and organize their own behaviors. The goal of a client-centered interview, then, is to create the type of interpersonal relationship most likely to enhance this self-change. In contrast, a behavioral interview is more likely to be based on the assumption that change occurs because of specific external influences and consequences. As a result, behavioral interviews are relatively structured because they are directed toward obtaining specific information that would help to design strategies to alter external conditions. In addition, different interviewing styles and strategies work well with some clients but may be relatively ineffective with others.

A useful distinction is between a diagnostic interview and one that is more informal and exploratory. The goal of a diagnostic interview is to develop a specific diagnosis, which usually was formerly based on the multiaxial *DSM-IV* model (see Othmer & Othmer, 1994; R. Rogers, 2001; Sommers-Flanagan & Sommers-Flanagan, 2013) but now is evolving to be based on the *DSM-5* (American Psychiatric Association [APA], 2013) taxonomy. Developing a diagnosis might follow a five-step process in which the clinician develops diagnostic clues, considers these in relation to diagnostic criteria, takes a psychiatric history, and, based on this information, develops a diagnosis with corresponding estimates of prognosis (Othmer & Othmer, 1994). Such an interview is likely to be directive with a careful consideration for inclusion and exclusion criteria of different disorders. It is most likely to occur in a psychiatric or general medical setting. In contrast, many practitioners do not believe in the value of formal diagnosis and, accordingly, do not pursue a formal *DSM-5* diagnosis. Even those who do value formal diagnosis may believe that the purpose of the clinical interview is to understand context, history, and interviewee's perspective, and the full assessment can work toward determining a formal diagnosis. These interviewers might be more concerned with areas such as a client's coping style, social supports, family dynamics, or the nature of the disability. As such, their interviews might be less directive and more flexible. Again, neither style is right or wrong, but one style may be appropriate and effective in one context (or client), whereas it is ineffective or inappropriate in another context.

Often interviewers wish to construct a semistructured interview format by listing in sequence the types of questions they would like to ask the person. To construct such a list, interviewers might consult Table 3.1 to note possibly relevant areas (note that this list is not exhaustive). Each of these areas might then be converted into specific questions, often starting with the most broad and general question and becoming

Table 3.1 Checklist for an Assessment Interview and Case History***Presenting Problem and Its History***

Description of the problem	Intensity and duration
Initial onset	Previous treatment
Changes in frequency	Attempts to solve
Antecedents/consequences	Formal treatment

Family Background

Socioeconomic level	Cultural background
Parents' occupations(s)	Parents' current health
Emotional/medical history	Family relationships
Married/separated/divorced	Urban/rural upbringing
Family constellation	

Personal History***Infancy***

Developmental milestones	Early medical history
Family atmosphere	Toilet training
Amount of contact with parents	

Early and middle childhood

Adjustment to school	Peer relationships
Academic achievement	Relationship with parents
Hobbies/activities/interests	Important life changes

Adolescence

All areas listed for early and middle childhood	Early dating
Presence of acting out (legal, drugs, sexual)	Reaction to puberty
	Childhood abuse

Early and middle adulthood

Career/occupation	Domestic violence
Interpersonal relationships	Medical/emotional history
Satisfaction with life goals	Relationship with parents
Hobbies/interests/activities	Economic stability
Romantic relationship/marriage	Substance abuse

Late adulthood

Medical history	Reaction to declining abilities
Ego integrity	Economic stability

Miscellaneous

Self-concept (like/dislike)	Somatic concerns (headaches, stomachaches, etc.)
Happiest/saddest memories	Events that create happiness/sadness
Earliest memory	Recurring/noteworthy dreams
Fears	

progressively more specific as needed. For example, the first few areas might be converted into this series of questions:

- “Tell me about the most important concerns that you have right now.”
- “How do these things affect you in your life?”
- “When did the difficulty first begin?”
- “How often does it occur?”
- “Have there been times when it has been better or worse?”
- “What happens after the behavior(s) occurs?”

Because clients vary regarding their personal characteristics (e.g., age, educational level, degree of cooperation) and type of presenting problem (e.g., childhood difficulties, legal problems, psychosis), interview questions necessarily need to vary from person to person. Furthermore, any series of questions should not be followed rigidly but with a certain degree of flexibility, to allow exploring unique but relevant areas that arise during the interview.

Good interviewing is difficult to define, partly because different theoretical perspectives exist regarding clinician–client interaction. Furthermore, clinicians achieve successful interviews not so much by what they do or say but by making sure they express the proper attitude. Whereas clinicians from alternative theoretical orientations might differ regarding areas such as their degree of directiveness or the type of information they should obtain, most agree that certain aspects of the relationship are essential (Patterson, 1989). These aspects include the interviewer’s expression of sincerity, acceptance, understanding, genuine interest, warmth, and a positive regard for the worth of the person. If clinicians do not demonstrate these qualities, they are unlikely to achieve the goals of the interview, no matter how these goals are defined.

Patient ratings of the quality of interviews have been found to be dependent on the extent to which interviewers can understand the patient’s emotions and detect emotional messages that are only partially expressed, particularly as these emotions are likely to be indirect and conveyed through nonverbal behaviors. Understanding a client’s emotional responses is especially relevant in clinical interviews that focus on a client’s personal difficulties. Typically, words are inadequate to accurately describe problem emotions, so interviewers must infer them from paraverbal or nonverbal expression. Reliance on nonverbal cues is highlighted by the assumption that nonverbal aspects of communication are a powerful method of conveying information. For example, eye contact can convey involvement; rigidity of posture might suggest client defensiveness; and hand movements often occur beyond the person’s conscious intent, suggesting nervousness, intensity, or relaxation. Mehrabian (1972) supported this perspective with his estimates that the message received is 55% dependent on facial expression, 38% dependent on tone, and only 7% dependent on the content of what is said.

Interviewers vary in the extent to which they take notes during the interview. Some argue that note taking during an interview might increase a client’s anxiety, raise questions regarding anonymity, increase the likelihood that a client will feel like an object under investigation, and create an unnatural atmosphere. In contrast, many interviewers counter these arguments by pointing out that a loss of rapport rarely

results solely from note taking during the interview, assuming, of course, that the interviewer still spends a sufficient amount of time attending to the client. Ongoing note taking is also likely to capture more details and result in less memory distortion than recording material after an interview has been completed. Thus, an intermediate amount of note taking during the interview is recommended. If the interview is audiotaped or videotaped, the reasons for this procedure need to be fully explained, along with the assurance of confidentiality and the procuring of signed consent. Although audiotape or videotape recording is often awkward at first, usually the interviewer and client quickly forget that it is occurring.

Interview Tactics

Numerous interview tactics and types of statements have been proposed and studied. These include the clarification statement, verbatim playback, probing, confrontation, understanding, active listening, reflection, feedback, summary statement, random probing, self-disclosure, perception checking, use of concrete examples, and therapeutic double binds. Additional relevant topics are the importance of eye contact, self-disclosure, active listening, and touch. These areas are beyond the scope of this chapter, but the interested reader is referred to excellent discussions by Cormier and Cormier (1998), Sommers-Flanagan and Sommers-Flanagan (2013), Sattler (2014), and Zuckerman (2005). The most relevant skills for interviewing do not come so much from memorizing interviewing tactics but from experiential practice and reviewing actual live or taped interview sessions. However, several important tactics of interviewing are described because they provide a general interviewing strategy.

Preliminaries

During the initial phase of the interview, practitioners need to ensure that they deal adequately with the next seven issues:

1. Organize the physical characteristics of the interview situation so that the room looks lived in but not untidy; utilize optimal lighting; and arrange seating so that the interviewer and client are neither too close nor too far and so that eye level is approximately equal.
2. Introduce themselves and indicate how they prefer to be addressed (Doctor, first name, etc.) and clarify how the client prefers to be addressed.
3. State the purpose of the interview, check the client's understanding of the process, and clarify any discrepancies between these two understandings.
4. Explain how the information derived from the interview will be used.
5. Describe the confidential nature of the information, the limits of confidentiality, and special issues related to confidentiality (e.g., how the information might be obtained and used by the legal system). Further, explain that the client has the right not to discuss any information he or she does not wish to disclose. If the information will be sent to other persons, obtain a signed release of information.
6. Explain the role and activities they would like the client to engage in, the instruments that are likely to be used in the assessment, and the total likely length

of time required. In some circumstances, this may be formalized into a written contract (Handelsman & Galvin, 1988).

7. Make sure that any fee arrangements have been clarified, including the hourly rate, total estimated cost, the amount the client versus a third party is likely to need to pay, and the interval between billing and the expected payment.

With the possible exception of fee arrangement (item 7), the preceding issues should be handled by a mental health practitioner rather than a secretary or receptionist. Covering these areas during the preliminary stages of the interview should reduce the likelihood of miscommunications and later difficulties.

Directive Versus Nondirective Interviews

The degree to which clinicians choose to be structured and directive during an interview depends on both theoretical and practical considerations. If time is limited, the interviewer will likely need to be direct and to the point. The interviewer will use a different approach for assessing a person who has been referred and will be returning to the referring person than for assessing a person before conducting therapy with him or her. An ambiguous, unstructured approach may make an extremely anxious person even more anxious, while a direct approach may prove more effective. A passive, withdrawn client also is likely to initially require a more direct question-and-answer style. As stated previously, a less structured style often encourages deeper client self-exploration, enables clinicians to observe the client's organizational abilities, and may result in greater rapport, flexibility, and sensitivity to the client's uniqueness.

Frequently, behavioral interviews are characterized as being structured and directed toward obtaining a comprehensive description of actual behaviors and relevant cognitions, attitudes, and beliefs (see Chapter 4). Behavioral interviewing is often contrasted with the more unstructured psychodynamic approach, which investigates underlying motivations and hidden dynamics and assesses information that may not be within the person's ordinary awareness. Typically, these approaches are perceived as competing and mutually exclusive. Haas, Hendin, and Singer (1987) pointed out that this either/or position is not only unnecessary but unproductive, because each style of interviewing provides different types of information that could potentially compensate for the other's weaknesses. Using both approaches might increase interview breadth and validity. Exploring multiple facets of the person may include direct behavioral data (public communication), self-description, and private symbolization (Leary, 1957). Each of these levels may be useful for different purposes, and the findings from each level might be quite different from one another.

Sequence of Interview Tactics

Most authors recommend that interviewers begin with open-ended questions and then, after observing the client's responses, use more direct questions to fill in gaps in their understanding (Harwood, Beutler, & Groth-Marnat, 2011; Othmer & Othmer, 2002; Sommers-Flanagan & Sommers-Flanagan, 2013). Although this sequence might begin with open-ended questions, it should typically lead to interviewer responses that are

intermediate in their level of directiveness, such as facilitating comments, requesting clarification, and possibly confronting the client with inconsistencies.

An important advantage of open-ended questions is that they require clients to comprehend, organize, and express themselves with little outside structure. This occasion is perhaps the only one in the assessment process that makes this requirement of clients, because most tests or structured interviews provide guidance in the form of specific, clear stimuli. When clients are asked open-ended questions, they will be most likely to express significant but unusual features about themselves. Verbal fluency, level of assertiveness, tone of voice, energy level, hesitations, and areas of anxiety can be noted. Hypotheses can be generated from these observations, and further questioning and testing can be used to test these hypotheses. In contrast to these advantages, open-ended questions can potentially provide an overabundance of detailed, vague, or tangential information.

Interviewer responses that show an intermediate level of directiveness are facilitation, clarification, empathy, and confrontation. Facilitation of comments maintains or encourages the flow of conversation. This might be accomplished verbally (“Tell me more . . .”; “Please continue . . .”) or nonverbally (eye contact, nodding). These requests for clarification might be used when clients indicate, perhaps through subtle cues, that they have not fully expressed something regarding the topic of discussion. Requests for clarification can bring into the open material that was only implied. In particular, greater clarification might be achieved by requesting the client to be highly specific, such as asking him or her to provide concrete examples (e.g., a typical day or a day that best illustrates the problem behavior). Empathic statements (“It must have been difficult for you”) can also facilitate client self-disclosure.

Sometimes interviewers may wish to confront, or at least comment on, inconsistencies in a client’s information or behavior. Carkhuff (1969) categorized the potential types of inconsistencies as being between what a person is versus what he or she wants to be, between what he or she is saying versus what he or she is doing, and between the person’s self-perception versus the interviewer’s experience of the person. A confrontation might also challenge the improbable content of what he or she is reporting (tall stories).

The purpose of confrontations during assessment is to obtain more in-depth information about the client. In contrast, therapeutic confrontations are used to encourage client self-exploration and behavior change. If a practitioner is using the initial interview and assessment as a prelude to therapy, this distinction is less important. However, a confrontational style can produce considerable anxiety, which should be created only if sufficient opportunity exists to work through the anxiety. Usually a client is most receptive to confrontations when they are posed either hypothetically as possibilities to consider or as curiosities on the part of the interviewer rather than as direct challenges. Confrontations also require a sufficient degree of rapport to be sustained; unless this rapport is present, confrontations may result in client defensiveness and a deterioration of the relationship.

Finally, direct, close-ended questions can be used to fill in gaps in what the client has reported. Thus, a continual flow can be formed between client-directed or client-organized responses and clinician-directed responses. This sequence, beginning with

open-ended questions, then moving to intermediately structured responses (facilitation, clarification, confrontation), and finally ending in directive questions, should not be rigid but should vary throughout the interview.

Comprehensiveness

The basic focus of an assessment interview should be to define the problem behavior (nature of the problem, severity, related affected areas) and its context (conditions that worsen or alleviate it, origins, antecedents, consequences). Interviewers might wish to use a checklist, such as the one in Table 3.1, to ensure they are covering the most relevant areas. In using such a checklist, the interviewer might begin with a general question, such as “How were you referred here?” or “What are some areas that concern you?” Observations and notes can then be made about the way the client organizes his or her responses, what he or she says, and the way he or she says it. The interviewer can use facilitating, clarifying, and confronting responses to obtain more information. Finally, the interviewer can review the checklist—for example, on family background—to see if all relevant areas were covered sufficiently. If some areas or aspects of areas were not covered, the interviewer might ask direct questions, such as “What was your father’s occupation?” or “When did your mother and father divorce?” The interviewer can then begin the same sequence for personal history related to infancy, middle childhood, and so on. Table 3.1 is not comprehensive but is intended as a general guide for most interview situations. If practitioners generally evaluate specific client types (e.g., child abuse, suicide, those with brain impairments), this checklist may need amending and/or be used as an adjunct to commercially available structured interviews, such as the Personality Disorder Examination (Loranger, 1988), Neuropsychological Status Examination (Schinka, 1983), or Lawrence Psychological-Forensic Examination (Lawrence, 1984).

Avoidance of “Why” Questions

It is best to avoid “why” questions because they can increase client defensiveness. A “why” question may sound accusatory or critical and thus forces the client to account for his or her behavior. In addition, clients may become intellectual in this situation, thereby separating themselves from their emotions. An alternative approach is to preface the question with either “What is your understanding of ...” or “How did it occur that ...” rather than “why?” These options are more likely to result in a description rather than a justification and to keep clients more centered on their emotions.

Nonverbal Behaviors

Interviewers should also be aware of their own as well as their clients’ nonverbal behaviors. In particular, interviewers might express their interest by maintaining eye contact, being facially responsive, and attending verbally and nonverbally, such as through occasionally leaning forward.

Concluding the Interview

Any interview is bound by time constraints. An interviewer might help to ensure observance of these constraints by alerting the client when only 5 or 10 minutes remain until

the arranged completion of the interview. Doing this allows the client or interviewer to focus on final relevant information. There should also be an opportunity for the client to ask any questions or provide comments. At the end of an interview or assessment session, the interviewer may summarize the main themes of the interview and, if appropriate, make any recommendations.

MENTAL STATUS EVALUATION

The mental status exam was originally modeled after the physical medical exam; just as the physical medical exam is designed to review the major organ systems, the mental status exam reviews the major systems of psychiatric functioning (appearance, cognitive function, insight, etc.). Since its introduction into American psychiatry by Adolf Meyer in 1902, it has become the mainstay of patient evaluation in most psychiatric settings. Most psychiatrists consider it as essential to their practice as the physical examination is in general medicine (Rodenhauser & Fornal, 1991).

A mental status examination can be used as part of a formal psychological assessment for a variety of reasons. A brief mental status examination might be appropriate before assessment to determine the appropriateness of more formal psychological testing. If, for example, a patient were unable to determine where he or she was and had significant memory impairments, testing with most instruments might be too difficult and could result in needless distress. Brief screenings might also be used to determine basic case management issues, such as hospitalization or placing a patient under close observation. A mental status examination can be used as part of an assessment using formal psychological tests. The “raw” data from the exam can be selectively integrated with general background information to present a coherent portrait of the person and assist in diagnosis.

Despite its popularity among psychiatrists, this form of interviewing is not typically used by psychologists, partly because many areas reviewed by the mental status exam are already covered during the assessment interview and through the interpretation of psychological test results. Many psychological tests cover these areas in a more precise, in-depth, objective, and validated manner with scores that can be compared to appropriate norms. A client’s appearance, affect, and mood are usually noted by attending to behavioral observations. A review of the history and nature of the problem is likely to pick up areas such as delusions, misinterpretations, and perceptual disorders (hallucinations). Likewise, interview data and psychological test results typically assess a client’s fund of knowledge, attention, insight, memory, abstract reasoning, and level of social judgment. However, the mental status examination reviews all of the preceding areas in a relatively brief, systematic manner. Furthermore, there are situations, such as intakes in an acute medical or psychiatric hospital, where insufficient time is available to evaluate the client with psychological tests.

Numerous sources in the psychiatric literature provide thorough guidelines for conducting a mental status exam (Crary & Johnson, 1981; Othmer & Othmer, 2002; Robinson, 2001; Sadock & Sadock, 2010; Sommers-Flanagan & Sommers-Flanagan, 2013), and R. Rogers (2001) has provided a review of the more structured mental status exams. This literature indicates that practitioners vary widely in how they

conduct mental status examinations. The most unstructured versions involve merely the clinician's use of the mental status examination as a set of general guidelines. The more structured versions range from comprehensive instruments that assess both general psychopathology and cognitive impairment to those that focus primarily on cognitive impairment. For example, the comprehensive North Carolina Mental Status Examination (Ruegg, Ekstrom, Evans, & Golden, 1990) includes 36 items that are rated on a 3-point scale (not present, slight or occasional, marked or repeated) to cover the important clinical dimensions of physical appearance, behavior, speech, thought processes, thought content, mood, affect, cognitive functioning, orientation, recent memory, immediate recall, and remote memory. Another similar comprehensive instrument is the Missouri Automated Mental Status Examination Checklist (Hedlund, Sletten, Evenson, Altman, & Cho, 1977), which requires the examiner to make ratings on nine areas of functioning: general appearance, motor behavior, speech and thought, mood and affect, other emotional reactions, thought content, sensorium, intellect, and insight and judgment. The checklist includes 119 possible ratings, but the examiner makes ratings in only those areas he or she judges to be relevant.

Despite extensive development, the more comprehensive mental status examinations have not gained wide acceptance. In contrast, the narrower structured mental status examinations that focus more exclusively on cognitive impairment are used quite extensively. One of the most popular has been the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975). It includes 11 items designed to assess orientation, registration, attention, calculation, and language. It has excellent interrater and test-retest reliabilities (usually well above .80), correlates with WAIS IQs (.78 for verbal IQ), and is sensitive to global and left-hemisphere deficits (but not right-hemisphere impairment; R. Rogers, 2001; Tombaugh, McDowell, Kristjansson, & Hubley, 1996). Clinicians who wish to develop knowledge and skills in conducting mental status examinations are encouraged to consult the preceding sources.

The following descriptions of the typical areas covered serve as a brief introduction to this form of MSE interviewing. The outline is organized around the categories recommended by Crary and Johnson (1981), and a checklist of relevant areas is included in Figure 3.1. Interviewers can answer the different areas on the checklist either during or after a mental status examination. The tabled information can then be used to answer relevant questions relating to the referral question, to help in diagnosis, or to add to other test data. Such a checklist is important because clinicians not using similar checklists have been found to often omit crucial information (Ruegg et al., 1990).

General Appearance, Behavior, and Relatedness

This area assesses material similar to that requested in the "behavioral observations" section of a psychological report (see Chapter 15). A client's clothing, posture, gestures, speech, personal care/hygiene, and any unusual physical features, such as physical disabilities, tics, or grimaces, are noted. Attention is given to the degree to which the client's behavior conforms to social expectations, but this is placed in the context of his or her culture and social position. Relatedness toward the evaluator is also an important factor to note. Additional important areas are facial expressions, eye contact, activity level, degree of cooperation, notable physical characteristics, and attentiveness.

Mental Status Evaluation

Appearance		Within Norm	Notable	Details
	Grooming			
	Motor Activity			
	Coordination/Gait			
	Notes on Appearance			

Relatedness	Within Norm	Notable	Details
	Cooperative	Hostile	
	Friendly	Guarded	
	Relaxed	Seductive	
	Good eye contact	Poor eye contact	
	Notes on Relatedness		

Speech/ Language		Within Norm	Notable		Details
	Receptive				
	Expressive		Quiet	Clutter/Stutter	
			Loud	Rapid	
			Slow	Pressured	
	Notes on Speech/Language				

Affect/Mood		Within Norm	Notable		Details
	Affect	Expressive	Flat	Anxious	
		Good Range	Constricted	Sad	
			Angry	Labile	
			Mood-Incongruent	Inappropriate to Situation	
	Mood	Euthymic	Elevated	Angry	
			Depressed		
	Notes on Affect/Mood				

Thought Process	Within Norm	Notable		Details
	Goal directed	Tangential	Flight of Ideas	
	Logical	Circumstantial	Slow	
	Abstract	Magical	Rapid	
	Reasoning	Concrete	Loose	
	Notes on Thought Process			

Figure 3.1 Format for mental status and history

Mental Status Evaluation

Thought Content		Present	Not Present	Details
	Hallucinations			
	Delusions			
	Depressive Ideation			
	Suicidality			
	Aggressiveness			
	Homicidality			
	Notes on Thought Content			

Memory		Intact	Impaired	Details
	Short-Term			
	Long-Term			
	Notes on Memory			

Attention/ Concentration	Within Norm		Notable	Details
	Notes on Attention/Concentration			

Alertness/ Orientation	Within Norm	Notable		Details
	Alert	Lethargic	Disoriented	
	Oriented	Hypervigilant		
	Notes on Alertness			

Judgment/ Planning		Within Norm	Notable	Details
	Judgment			
	Impulse Control			
	Notes on Judgment/Planning			

Insight	Within Norm	Notable		Details
	Notes on Insight			

Figure 3.1 (Continued)

Is the client friendly, hostile, seductive, or indifferent? Do any bizarre behaviors or significant events occur during the interview? In particular, speech might be fast or slow, loud or soft, or include a number of additional unusual features. Figure 3.1 includes a checklist of relevant areas of behavior, appearance, and relatedness.

Speech and Language

Clients' speech and language are often proxies for their thought processes, as they relate to the primary mode of communicating thoughts to the outside world. They help clinicians determine the possibility of poor or exceptional cognitive functioning, focus, and confusion and possible thought disorder. Additionally, speech and language often highlight interpersonal characteristics, such as shyness, anxiety interacting with others, and aggressiveness. Clinicians should evaluate in general how well individuals understand language, as evidenced by responding appropriately to directions and conversations (known as *receptive* language). *Expressive* language, in contrast, relates to the client's actual speech and use of language. *Speech* relates to the quality of speaking, such as quiet, loud, rapid, slow, and so on. *Language* relates to the words used, including having difficulty with word finding, using complex and appropriate vocabulary, or misusing words often.

Feeling (Affect and Mood)

A client's *mood* refers to the dominant emotion reported during the interview, whereas *affect* refers to the client's outwardly projected range of emotions. Information related to affect is inferred from the content of the client's speech, facial expressions, and body movements. The type of affect can be judged according to variables such as its depth, intensity, duration, and appropriateness. The client might be cold or warm, distant or close, labile, or, as is characteristic of schizophrenia, his or her affect might be blunted or flattened. The client's mood might also be euphoric, hostile, anxious, or depressed, and an examiner should note the level of congruence between mood and affect.

Perception and Thinking

Perception

Different clients perceive themselves and their world in a wide variety of ways. It can be diagnostically important to note whether there are any illusions or hallucinations. For example, the presence of auditory hallucinations is most characteristic of those with schizophrenia, whereas vivid visual hallucinations are more characteristic of persons with organic brain syndromes.

Intellectual Functioning

Any assessment of higher intellectual functioning needs to be made in the context of a client's educational level, socioeconomic status, and familiarity and identification with a particular culture. If a low level of intellectual functioning is consistent with a general pattern of poor academic and occupational achievement, a diagnosis of intellectual

disability might be supported. However, if a person performs poorly on tests of intellectual functioning and yet has a good history of achievement, organicity might be suspected.

Intellectual functioning typically involves reading and writing comprehension, general fund of knowledge, ability to do arithmetic, and the degree to which the client can interpret abstract language, such as proverbs. Throughout the assessment, clinicians typically note the degree to which the client's thoughts and expressions are articulate versus incoherent. Sometimes clinicians might combine assessments of intellectual functioning with some short, formal tests such as the Bender, an aphasia screening test, or even portions of the WAIS or WISC.

Orientation

The ability of clients to be oriented can vary in the degree to which they know who they are (person), where they are (place), and when current and past events have occurred or are occurring (time). Clinical observation indicates the most frequent type of disorientation is for time; disorientation for place and person occurs less frequently. When disorientation does occur for place, and especially for person, the condition is likely relatively severe. Disorientation is most consistent with organic conditions. If a person is oriented in all three spheres, this is frequently abbreviated as "oriented X3."

Related to the orientation of clients is their *sensorium*, which refers to how intact their physical sensory processes are to receiving and integrating information. *Sensorium* might refer to hearing, smell, vision, and touch and might range from being clouded to clear. Can the client attend to and concentrate on the outside world, or are these processes interrupted? The client might experience unusual smells, hear voices, or have the sense that his or her skin is tingling. *Sensorium* can also refer to the client's level of consciousness, which may vary from hyperarousal and excitement to drowsiness and confusion. Disorders of a client's sensorium often reflect organic conditions but may also be consistent with psychosis.

Memory, Attention, and Concentration

Because memory acquisition and retrieval require attention and concentration, these three functions are frequently considered together. Long-term memory is often assessed by requesting information regarding the client's general fund of information (e.g., important dates, major cities in a country, three major heads of state since 1900). Some clinicians include the Information or Digit Span subtests from the WAIS/WISC or other formal tests of a similar nature. Recall of a sentence or paragraph might be used to assess short-term memory for longer, more verbally meaningful information. In addition, clients' long-term memory might be evaluated by measuring recall of major life events, and the accuracy of their recall can be compared with objective records of these events (e.g., year graduated from high school, date of marriage). It is often useful to record any significant distortions of selective recall in relation to life events, as well as to note the client's attitudes toward his or her memory.

Short-term memory might be assessed either by requesting that clients recall recent events (e.g., most recent meal, how they got to the appointment) or by having them repeat digits forward and backward. Again, the WAIS/WISC Digit Span subtest, or

a similar version of it, might be used. Serial sevens (counting forward by adding 7 each time) can be used to assess how distractible or focused they are. Persons who are anxious and preoccupied have a difficult time with serial sevens as well as with repeating digits forward and, especially, backward.

Insight and Judgment

Clients vary in their ability to interpret the meaning and impact of their behavior on others. They also vary widely in their ability to provide for themselves, evaluate risks, and make plans. Adequate insight and judgment involve developing and testing hypotheses regarding their own behavior and the behavior of others. Clients also need to be assessed to determine why they believe they were referred for evaluation and, in a wider context, their attitudes toward their difficulties. How do they relate their past history to current difficulties, and how do they explain these difficulties? Where do they place the blame for their difficulties? Based on their insights, how effectively can they solve problems and make decisions?

Thought Content

A client's speech can often be considered a reflection of his or her thoughts. The client's speech may be coherent, spontaneous, and comprehensible or may contain unusual features. It may be slow or fast, be characterized by sudden silences, or be loud or unusually soft. Is the client frank or evasive, open or defensive, assertive or passive, irritable, abusive, or sarcastic? Consideration of a person's thoughts is often divided into thought content and thought processes. Thought content such as delusions might suggest a psychotic condition, but delusions may also be consistent with certain organic disorders, such as dementia or chronic amphetamine use. The presence of compulsions or obsessions should be followed up with an assessment of the client's degree of insight into the appropriateness of these thoughts and behaviors. Thought processes such as the presence of rapid changes in topics might reflect flighty ideas. The client might also have difficulty producing a sufficient number of ideas, include an excessive number of irrelevant associations, or ramble aimlessly.

INTERPRETING INTERVIEW DATA

Interpreting and integrating interview data into the psychological report inevitably involves clinical judgment. Even with the use of structured interviews, the clinician still must determine which information to include or exclude. Thus, all the potential cautions associated with clinical judgment need to be considered (see Chapter 1). Caution is particularly important because life decisions and the success of later treatment may depend on conclusions and recommendations described in the report.

Several general principles can be used to interpret interview data. The interview is the primary instrument that clinicians use to develop tentative hypotheses regarding their clients. Thus, interview data can be evaluated by determining whether these hypotheses are supported by information outside the interview. Interview data that are supported by test scores can be given greater emphasis in the final report if they are

relevant to the referral question. Even material that is highly supported throughout different phases of the interview process should not be included unless it relates directly to the purpose of the referral.

There is a continuum in handling interview information that varies according to the extent the information will be interpreted. On one hand, the information might be merely reorganized into a chronological history of the person's life. This method would emphasize repeating the information in as objective and accurate a manner as possible. Typically this is done in the history section of a psychological report. On the other hand, interview data can be considered raw data to be interpreted. It is thus similar to the data from formal psychological tests. It might, therefore, be used to make inferences related to a client's personality, coping style, or mood and affect.

One method of organizing interview information is to use the information to develop a coherent narrative of the person's life. For example, describing how early family patterns resulted in emotionally sensitive areas ("scar" tissue) can be used to help explain current symptom patterns and difficulties in interpersonal relationships. A different sort of history might trace how interest in a vocation was first begun (early childhood daydreams regarding occupations) and how this progressed and developed as the person matured. Another person might present difficulties related to authority figures. Specific details relating to these difficulties might emerge, such as the client feeling like a martyr and eventually inappropriately expressing extreme anger toward authority figure(s). A careful review of the client's history might reveal how he or she becomes involved in these recurring relationships and how he or she typically attempts to resolve them. Persons who are frequently depressed might distance themselves from others by their behavior and then be confused about why relationships seem to be difficult. Often these themes emerge during a carefully conducted interview, yet aspects of the themes (or the entire themes themselves) are not apparent to the interviewee.

Interview data might also be organized around various domains (see further discussion in Chapter 15). A grid can be used to organize these domains. The various domains might be listed on the left side of the grid with the top of the grid listing the sources of data (of which the interview might be one of a variety of sources of information; see Table 15.2 in Chapter 15). Domains might include mood and affect, cognitions, level of resistance, interpersonal patterns, or coping style. This approach treats interview data in much the same manner as data from psychological tests.

There is no one strategy for sensitizing interviewers to the types and patterns of recurring themes they may encounter during interviews. Inevitably, clinical judgment is a significant factor. The accuracy and types of judgment depend on the theoretical perspective of the interviewer, knowledge regarding the particular difficulty the interviewer is investigating, past experience, types of questions asked, and purpose of the interview.

STRUCTURED INTERVIEWS

Standardized psychological tests and structured interviews were developed to reduce the problems associated with open-ended interviews. They serve both to structure the stimuli presented to the person and to reduce the (potentially biased) role of clinical

judgment. Because structured interviews generate objective ratings on consistent areas, they have the advantage of making possible comparisons between one case or population and the next. Typically, these interviews vary in their degree of structure, the relative expertise required to administer them, and the extent to which they serve as screening procedures designed for global measurement or as tools used to obtain specific diagnoses.

Before structured interviews could be developed, clear, specific criteria had to be created relating to symptom patterns and diagnoses. Developing these clear, specific criteria ideally helped to reduce the amount of error caused by vague guidelines for exclusion or inclusion in different categories (*criterion variance*). These criteria then needed to be incorporated into the interview format and interview questions. *Information variance* refers to the variability in amount and type of information derived from interviews with clients. In most unstructured interviews, information variance is caused by the wide differences in content and phrasing because of factors such as the theoretical orientation or style of the interviewer. Structured interviews correct for this by utilizing the same or similar questions for each client.

The first popular system of specific criterion-based diagnosis was developed by Feighner et al. (1972) and provided clear, behaviorally oriented descriptions of 16 psychiatric disorders based on the *DSM-II* (APA, 1968). Clinicians using the Feighner criteria were found to have an immediate and marked increase in interrater diagnostic reliability. The descriptions of and relevant research on the Feighner criteria were published in Woodruff, Goodwin, and Guze's (1974) book, *Psychiatric Diagnosis*. Several interviews, such as the Renard Diagnostic Interview (Helzer et al., 1981), incorporated the Feighner criteria. Spitzer, Endicott, and Robins (1978) further altered and elaborated the Feighner criteria to develop the Research Diagnostic Criteria (RDC). Simultaneous with the development of the RDC, Endicott and Spitzer (1978) developed the SADS, which was based on the new RDC. When new versions of the *DSM* were published (APA, 1980, 1987, 1994, 2000, 2013), revisions of previous interviews typically incorporated the most recent *DSM* criteria, along with elements of the Feighner criteria and/or the RDC.

As noted earlier, the reliability of structured interviews has been found to vary depending on the specificity or precision of the rating or diagnosis. Whereas the highest reliabilities have been found for global assessment (presence/absence of psychopathology), much lower reliabilities have generally been found for the assessment of specific types of behaviors or syndromes. Likewise, high reliabilities have been found for overt behaviors, but reliability has been less satisfactory for more covert aspects of the person, such as obsessions, fears, and worries. Reliability also tends to be lower when clinicians are asked to attempt exact estimates of behavioral frequencies and for inferences of multifaceted aspects of the person derived from complex clinical judgments.

Most early studies on validity were based on item content (content validity) or degree of accuracy in distinguishing between broad areas of psychopathology (psychiatric/nonpsychiatric). More recent trends have attempted to assess the accuracy of far more specific areas. However, most validity studies have suffered from an absence of clear, commonly agreed-upon criteria. Although structured interviews were attempts to improve on previous, imperfect instruments (unstructured interviews, standardized tests), the structured interviews themselves could not be compared with anything

better. For example, the “procedural validity” strategy is based on comparing lay interviewers’ diagnoses with diagnoses derived from trained psychiatrists. Although the psychiatrist’s diagnosis may be better than the layperson’s, diagnoses by trained psychiatrists still cannot be said to be an ultimate, objective, and completely accurate standard. Furthermore, there is confusion about whether actual validity is being measured (which would assume psychiatrists’ diagnoses are the true, accurate ones) or merely a version of interrater agreement. At the core of this issue is the very nature of how diagnosis is defined and the degree to which it is actually helpful in treatment (see Beutler & Malik, 2002; Widiger & Clark, 2000).

Future studies need to involve aspects of what has previously been discussed as construct validity. The focus on construct validity means looking more carefully at structured interviews in relationship to etiology, course, prognosis, and treatment utility relating to areas such as the appropriate selection of treatments and the likelihood of favorable responses to these treatments. Validity studies also need to look at the interaction between and implications of multiple criterion measures, including behavioral assessment, checklists, rating scales, self-report inventories, biochemical indices, and neuropathological alterations.

Since the mid-1970s, there has been a proliferation of structured interviews for a wide range of areas. Clinicians working in specific areas often select structured interviews directed toward diagnosing the disorders they are most likely to encounter. For example, some situations might benefit from using the Anxiety Disorders Interview Schedule–IV (DiNardo, Brown, & Barlow, 1994) to make clear distinctions between anxiety disorders and substance abuse and between psychosis and major affective disorders. Other contexts might be best served by the Eating Disorder Examination (EDE; Z. Cooper & Fairburn, 1987) or the Structured Interview for *DSM-IV* Dissociative Disorders (SCID-D; Steinberg, 1993). Three categories of structured interviews with representative frequently used instruments are included in Table 3.2 and have been extensively reviewed in R. Rogers’s (2001) *Handbook of Diagnostic and Structured Interviewing*. One consideration in selecting these instruments is that, because most structured interviews undergo continuous revisions, the most up-to-date research should be consulted to ensure that practitioners obtain the most recently revised versions. The next section provides an overview of structured interviews that are used most frequently and are the most extensively researched.

Structured Clinical Interview for the *DSM*

The SCID (First, Spitzer, Gibbon, & Williams, 1996, 1997; Spitzer et al., 1987) is the most frequently used structured interview (see description and updates at www.scid4.org and www.appi.org/pages/scid-5.aspx). It is a clinician-administered, comprehensive broad-spectrum instrument that adheres closely to the *DSM* decision trees for psychiatric diagnosis. A certain degree of flexibility is built in so that administration can be tailored to different populations and contexts. Thus, slightly different forms are used for psychiatric patients (SCID–In/Patient), outpatients (SCID–Out/Patients), and nonpatients (SCID–Non/Patients). Criticisms that the early version of the SCID had sacrificed clinical information so that it would be more user-friendly for clinicians resulted in a revision that emphasized a clear, easy-to-use version for clinical contexts

Table 3.2 Frequently Used Structured Interviews by Categories

I. Assessment of clinical disorders
Schedule of Affective Disorders and Schizophrenia (SADS) and Schedule of Affective Disorders and Schizophrenia for School-Age Children (K-SADS)
Diagnostic Interview Schedule (DIS) and Diagnostic Interview for Children (DISC)
Structured Clinical Interview for DSM-IV (SCID)
Diagnostic Interview for Children and Adolescents (DICA)
II. Assessment of personality disorders
Structured Interview for DSM-IV Personality Disorders (SIDP)
Personality Disorder Examination (PDE)
Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II)
III. Focused structured interviews
Anxiety Disorders Interview Schedule (ADIS)
Diagnostic Interview for Borderlines (DIB)
Psychopathy Checklist (PCL)
Structured Interview for DSM-IV-Dissociative Disorders (SCID-D)
Structured Interview of Reported Symptoms (SIRS)
Psychosocial Pain Inventory (PSPI)
Comprehensive Drinker Profile (CDP)
Eating Disorder Examination (EDE)
Structured Interview of Sleep Disorders (SIS-D)
Substance Use Disorders Diagnostic Schedule (SUDDS)

(the SCID-Clinical Version or SCID-CV; First et al., 1997) and a longer, more in-depth version for research (SCID-I or SCID-Research Version; First, Spitzer et al., 1996). A new version aligned with the *DSM-5* is now available as well (SCID-5; First, Williams, Karg, & Spitzer, 2015). Whereas these versions of the SCID are directed toward what used to be known as Axis I diagnoses, a separate version has been developed for the diagnosis of personality disorders (SCID-II; Spitzer, Williams, Gibbon, & First, 1990). A further variation, the SCID-D-Revised (Steinberg, 1993), was developed (though not by the team who developed the SCID) using *DSM-IV* criteria for the assessment of dissociative disorders. The SCID and its variations include several open-ended questions as well as a skip structure, which enables the interviewer to branch into new areas depending on the client's previous responses. Because clinical judgment is essential throughout the interview, the SCID should be administered only by trained professionals. To increase incremental validity, the authors encourage the inclusion of relevant additional data in making final diagnostic decisions.

The SCID, along with its variations, is the most comprehensive structured interview available. As a result, administration time can be considerable, even with the built-in screening questions and skip structure. Many individual clinicians and treatment sites deal with this problem by primarily administering the modules they are

most concerned with. For example, a treatment center specializing in substance abuse might administer the module for Psychoactive Substance Use Disorders along with the SCID-II when the comorbidity of personality disorders is suspected. Administration time might also be reduced by administering the computerized mini-SCID (First, Gibbon, Williams, & Spitzer, 1996), which has been designed to screen for possible (formerly) Axis I disorders. In addition, a computerized SCID-II (AutoSCID-II; First, Gibbon et al., 1996) that can potentially reduce clinician time is available. Although it can be administered by telephone, this procedure is discouraged, given the poor agreement between telephone and face-to-face diagnoses (Cacciola, Alterman, Rutherford, McKay, & May, 1999).

The reliability studies have resulted in overall moderate, but quite variable, test-retest and interrater reliabilities (First & Gibbon, 2004). For example, interrater agreement using the SCID-II for common diagnostic categories ranges between .40 and .86 with a mean of .59 (First, Spitzer, Gibbon, & Williams, 1995). Riskind, Beck, Berchick, Brown, and Steer (1987) found that several difficult-to-distinguish diagnostic categories had relatively good levels of interrater agreement. These included generalized anxiety disorders (.79, 86% agreement), depressive disorders (.72, 82% agreement; Riskind et al., 1987), panic disorders ($k = .86$), and major depression ($k = .81$; J. Reich & Noyes, 1987). Test-retest reliabilities over a 2-week interval for psychiatric patients was fair to good (overall weighted kappas = .61) but poor for nonpatients (overall weighted kappas = .37; J. B. Williams et al., 1992).

For the most part, validity studies of the SCID have assumed that *DSM-IV* diagnoses are the benchmark for making comparisons of diagnostic accuracy. Thus, “procedural validity” has often been assumed since the SCID has closely paralleled the diagnostic criteria derived from the *DSM* (R. Rogers, 2001). A representative validity study found good agreement ($k = .83$) between interviewer ratings and cross ratings of interviewer videotapes by two senior psychiatrists (Maziade et al., 1992). Other studies have found considerable diagnostic overlap within (formerly) Axis I disorders and between (formerly) Axis I and personality disorders (Alnacs & Torgerson, 1989; Brawman-Mintzer et al., 1993). However, evaluating the meaning of this overlap is difficult because the extent to which it is caused by instrument error versus true comorbidity (i.e., the frequent co-occurrence of anxiety and depression) is difficult to determine. In contrast to these mostly favorable studies, a number of studies have found generally poor agreement between SCID and clinician-based diagnosis (Shear et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995). In summary, the strength of the SCID is its impressive breadth of coverage, use of modules targeted toward specific areas, and close parallel with the *DSM*. Its weaknesses are its wide variation in reliability and its need for further validity studies, particularly studies that relate it to other diagnostic measures.

Schedule for Affective Disorders and Schizophrenia

The SADS (Endicott & Spitzer, 1978) is a clinician-administered, extensive, semistructured interview that has been one of the most widely used structured interviews for clinical research purposes. Although it was originally designed for differential diagnosis between affective disorders and schizophrenia, it has evolved to include a much

wider range of symptoms and allows the interviewer to consider many different diagnostic categories. A wide range of disorders is considered within the SADS, but its primary strength lies in obtaining fine detail regarding different subtypes of affective disorders and schizophrenia (R. Rogers, Jackson, & Cashiel, 2004). The interview rates clients on six gradations of impairment from which diagnoses are reached using the clear, objective categories derived from Spitzer et al.'s (1978) RDC. The SADS is divided into adult versions for current symptoms, occurrence of lifetime symptoms, and degree of change. There is a further version for the assessment of children's difficulties (K-SADS or Kiddie-SADS). Two modifications for the SADS have been the inclusion of anxiety disorders (SADS-LA; Fyer, Endicott, Manuzza, & Klein, 1985, 1995) and eating disorders (EAT-SADS-L; Herzog, Keller, Sacks, Yeh, & Lavori, 1992).

Adult Version

The adult version of the SADS (Endicott & Spitzer, 1978) is designed to be administered in two different parts, the first focusing on the client's current illness and the second on past episodes. This division roughly corresponds with the three different versions of the SADS. The first is the regular version (SADS), the second is the lifetime version (SADS-L, which is actually the second half of the SADS), and the third is the SADS-C, which measures changes in the client. The SADS-L is directed toward diagnosing the possible presence of psychiatric disturbance throughout the person's life. The SADS and SADS-L are used most extensively. Because the questions in the SADS are directed toward current symptoms and those symptoms experienced 1 week before administration, it is most appropriate for administration when the client is having current difficulties. In contrast, the SADS-L is most appropriate when there is no current, acute illness. To make accurate ratings, interviewers are allowed to use a wide range of sources (client's family, medical records) and ask a number of different questions. Final ratings are made on a 6-point Likert-type scale. Administration involves more than 200 items and takes from 1.5 to 2 hours and should be conducted only by a psychiatrist, psychologist, or psychiatric social worker. The end product is eight summary scales:

1. Mood and ideation
2. Endogenous features
3. Depressive-associated features
4. Suicidal ideation and behavior
5. Anxiety
6. Manic syndrome
7. Delusions-hallucinations
8. Formal thought disorder

Interrater reliabilities for the specific diagnostic categories have been found to be quite high, with the exception of the Formal Thought Disorder Scale (Endicott & Spitzer, 1978). The low reliability of this scale may have been because few of the patients in the Endicott and Spitzer sample showed clear patterns of disordered thoughts, which

resulted in high variability for the ratings. Test-retest reliabilities were likewise good, ranging from .88 for manic disorders to .52 for chronic and intermittent depressive disorder (Spiker & Ehler, 1984). The exception was a low reliability for schizoaffective, depressed (.24), but this was likely due to the small number of patients included in this category, which resulted in limited variance. Using a different and possibly more appropriate statistical method, reliability increased to .84. Overall, the SADS has demonstrated excellent reliability, particularly for interrater and test-retest reliabilities related to current episodes of psychiatric disturbance.

Validity studies have been encouraging because expected relationships have been found between SADS scores and external measures of depression, anxiety, and psychosis. For example, M. H. Johnson, Margo, and Stern (1986) found that relevant SADS measures could effectively discriminate between patients with depression and paranoid and nonparanoid schizophrenia. In addition, the SADS depression measures effectively rated the relative severity of a patient's depression. For example, Coryell et al. (1994) found clear consistency between different levels of depression. The authors suggest that incremental validity might be increased by having clients referred for a medical examination to screen out physical difficulties that might be resulting in central nervous system dysfunction. The authors also recommend that interviewers try to increase validity by always including the best available information (family history, structured tests, other rating schedules) before making final ratings. The SADS has been used to predict the clinical features, course, and outcome of various disorders, including major depressive disorder (Coryell et al., 1994), schizophrenia (Stompe, Ortwein-Swoboda, Strobl, & Friedman, 2000), and bipolar disorder (Vieta et al., 2000). A number of studies have also effectively used the SADS to detect family patterns of schizophrenia (Stompe et al., 2000) and obsessive-compulsive disorders (Bienvenu et al., 2000).

Child Version

The SADS for School-Age Children (Kiddie-SADS-P, K-SADS-P; Ambrosini, 2000; Puig-Antich & Chambers, 1978) is a semistructured interview developed for children between ages 6 and 18. The K-SADS has come out in versions to be used in epidemiological research (K-SADS-E), to assess present and lifetime psychopathology (K-SADS-P/L), and to assess current levels of symptomology (K-SADS-P). Although much of the K-SADS is based on research with major depressive disorders of prepubertal children, it also covers a wide range of other disorders, such as phobias, conduct disorders, obsessive-compulsive disorders, and separation anxiety.

The interview should be administered by a professional clinician who has been trained in the use of the K-SADS and is familiar with *DSM* criteria. All versions are administered to both the parent and the child. Any discrepancies between the two sources of information are clarified before final ratings are made. Total administration time is approximately 1.5 hours per informant (3 hours total). The first phase is a 15- to 20-minute unstructured interview in which rapport is developed as well as an overview of relevant aspects of history, including the frequency and duration of presenting symptoms, their onset, and whether the parents have sought previous treatment. This interview is followed by structured questions regarding symptoms, which are rated on a Likert scale, with 1 representing "not at all" and 7 indicating

that they are “extreme.” A skip structure is built into the format so that interviewers can omit irrelevant questions. Interviewers are allowed to use their judgment regarding the wording and the type and number of questions. Finally, ratings are made regarding behavioral observations (appearance, attention, affect). Interviewers are also asked to rate the completeness and reliability of the interview and to make a global assessment of pathology (degree of symptomatology and level of impairment).

Test-retest and interrater reliability for the K-SADS has been good with a general trend for each version to have improved reliabilities. Ambrosini (2000), for example, reported that the K-SADS-P/L had test-retest reliabilities ranging from 1.00 (lifetime occurrence of major depression) to .55 (for lifetime occurrence for attention-deficit/hyperactivity disorder). However, overall reliabilities have been lower for the K-SADS (and K-SADS-III-R) than for the adult SADS, although this is to be expected given the relative changeability and less well-developed language skills found with children (Ambrosini, Metz, Prabucki, & Lee, 1989; Chambers et al., 1985). Validity studies indicate that relevant K-SADS measures correlated highly with diagnoses for conduct disorders, schizophrenia, and depression (Apter, Bleich, Plutchik, Mendelsohn, & Tyrano, 1988). Additional expected correlations have been found between SADS measures and ratings of adolescent mood (Costello, Benjamin, Angold, & Silver, 1991) and the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983; Ambrosini, 2000). Finally, follow-up studies on adolescents diagnosed with disorders (e.g., depression) have found a continued risk for later affective difficulties (Lewinsohn, Rohde, Klein, & Seeley, 1999).

Collectively, the different versions of the SADS provide a thorough, well-organized interview with unparalleled coverage of the subtypes and gradations of the severity of mood disorders. The SADS has also been well accepted in research and clinical settings. It has strong interrater reliability, provides good ratings of symptom severity, measures associated symptoms, includes guidelines for possible malingering, and has strong evidence of convergent validity (see R. Rogers, 2001; R. Rogers et al., 2004). In contrast, its weaknesses include a relatively narrow band of diagnosis compared with some of the other available instruments, such as the SCID or DIS. In addition, the diagnoses are based on RDC rather than the more recent *DSM-IV-TR* or *DSM-5* criteria. This criticism is somewhat moderated, however, by many of the RDC and *DSM* criteria being nearly the same, especially for childhood disorders. Finally, administration and interpretation of the SADS require extensive training (usually 1 week) as well as a good working knowledge of differences between the SADS/RDC and *DSM* criteria.

Diagnostic Interview Schedule

In contrast to the SADS, which is semistructured and requires administration by trained professionals, the DIS (Robins, Helzer, Croughan, & Ratcliff, 1981) is highly structured and was designed specifically by the National Institute of Mental Health (Division of Biometry and Epidemiology) to be administered by nonprofessional interviewers for epidemiological studies (see Helzer & Robins, 1988). It has been updated for the *DSM-III-R* (Robins et al., 1989) and the *DSM-IV* (Robins, Cottler, Bucholz, & Compton, 1996) but has yet to be updated for the *DSM-5*. The latest version (DIS-IV) includes 19 modules with more than 30 diagnoses, including one

personality disorder diagnosis (antisocial personality). This modular format allows for tailoring various portions of the DIS-IV to the interests of the researcher or clinician. However, clinical judgment is reduced to a minimum by using verbatim wording, specific guidelines, a clear flow from one question to the next, and simple yes-no answers. Thus, the DIS is far more economical to administer than the SADS. Total administration time is 60 to 90 minutes. Studies have generally indicated that results are comparable between trained clinicians and nonprofessional interviewers (Helzer, Spitznagel, & McEvoy, 1987).

Adult Version

The original version of the DIS was derived from the format of the earlier Renard Diagnostic Interview. However, diagnosis for the DIS-IV is based exclusively on *DSM-IV* criteria. Initially, questions are directed toward obtaining information regarding the client's life, and information is also requested regarding more current symptoms based on the past 2 weeks, past month, past 6 months, and past year. Specific probe questions distinguish whether a symptom is clinically significant. A total of 470 potential clinical ratings are made and organized around 24 major categories. Administration time is approximately 60 to 90 minutes.

Computerized administration and scoring programs are available that can generate *DSM-IV*-based diagnoses. However, computer-based diagnoses on early versions of the DIS were found to generate an average of 5.5 possible diagnoses compared with an average of 2.6 for nonstructured interviews (Wyndowe, 1987). Patient acceptance for the computer administration has been found to be high, although the average administration time is somewhat longer than the clinician-interviewed version.

Studies of the reliability and validity of the DIS have been both variable and controversial. Although much of this research was done on pre-DIS-IV versions, the similarity of format and content between the DIS and DIS-IV suggests that much of this earlier research is pertinent. The comparability of diagnosis by professionals and nonprofessionals using the DIS has generally been supported. This finding suggests that nonprofessionals can effectively use the DIS to help gather data for large epidemiological studies. For example, Robins et al. (1981) found diagnostic agreement between psychiatrists and nonprofessional interviewers to be .69. The sensitivity (percentage of interviewees correctly identified) of the DIS varied according to type of diagnosis but had a mean of 75%, with a mean specificity (percentage of noncases correctly identified) of 94%. More recent studies have similarly concluded that the specificity is stronger than its sensitivity (Eaton, Neufeld, Chen, & Cai, 2000; J. M. Murphy, Monson, Laird, Sobol, & Leighton, 2000). However, data on sensitivity and specificity were based on using psychiatrists' diagnoses as the true index of diagnostic accuracy. The difficulties in considering psychiatrists' ratings as the truly accurate or "gold standard" criterion for validity have already been noted; therefore, it is probably best to consider the preceding data on sensitivity and specificity as forms of interrater agreement rather than concurrent validity. In contrast to this study, Vandiver and Sheer (1991) found somewhat marginal median test-retest reliabilities, ranging between .37 and .46.

Although many of the DIS ratings between professional and lay interviewers were equivalent, Helzer et al. (1985) found that, when compared with psychiatrists, nonprofessional interviewers tended to overdiagnose major depression. In contrast to Helzer et al. (1987), Folstein et al. (1985) did not find a sufficiently high rate of

agreement between diagnoses by a panel of psychiatrists and diagnoses by the DIS to warrant its use in epidemiological studies. Specifically, it was found that the DIS generated more cases of depression and schizophrenia and fewer cases of alcoholism and antisocial personality (Cooney, Kadden, & Litt, 1990; Folstein et al., 1985). Eaton et al. (2000) noted that false-negative diagnoses for many cases could be attributed mainly to failure by patients to report symptoms based on life crises or medical conditions. In contrast, the DIS has been found to be comparable with other commonly used psychiatric rating devices, such as the Psychiatric Diagnostic Interview (Folstein et al., 1985; R. Weller et al., 1985). However, both diagnostic strategies may contain inaccuracies, and it is difficult to tell in which areas these inaccuracies occurred (R. Weller et al., 1985). The DIS has had the greatest difficulty accurately diagnosing borderline conditions and patients in remission, but this is to be expected because these are the most problematic diagnoses for many other assessment strategies (Robins & Helzer, 1994). In contrast, Swartz et al. (1989) were able to find quite respectable sensitivities (85.7%) and specificities (86.2%) for borderline conditions using a DIS borderline index.

Child Version

The Diagnostic Interview Schedule for Children (DISC; Costello, Edelbrock, Duncan, & Kalas, 1984; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) is similar to the adult version in that it is highly structured and designed for nonprofessional interviewers. It differs in that it is designed to be given as both a child interview (DISC-C) and parent interview (DISC-P). There have also been versions designed for teachers (Teacher DISC), screening (DISC Predictive Scales), young adults (Young Adult DISC), and administrations that can be given by computer or audio recording (Lucas et al., 2001; Shaffer et al., 2000). Ratings are coded as 0 (not true), 1 (somewhat true), or 2 (very often true). *DSM-IV* (and now *DSM-5*) diagnoses are generated based on the combined ratings for the child and parent interviews. Some of the more problematic diagnoses (Autism, Pervasive Developmental Disorder, Pica) are based on an interview with the parent only. The entire interview takes an average of 70 minutes per informant and 90 to 120 minutes per patient, but an explicit skip structure can enable some interviews to be somewhat shorter. The most recent modification of the DISC (DISC-IV; Robins et al., 1996; Shaffer et al., 2000) was designed to be compatible with *DSM-IV* and *ICD-10* criteria. The DISC-IV comprises six modules, each of which represents the major diagnostic clusters (Anxiety, Mood, Disruptive, Substance Use, Schizophrenia, Miscellaneous).

DISC test-retest reliability (1-year interval) for *DSM-IV* diagnoses in a clinical sample was good to adequate with parent ratings having higher reliabilities (.54–.79) than child interviews (.25–.92; Shaffer et al., 2000). However, test-retest reliabilities for a community sample were generally quite poor for child interviews (.27–.64), although adequate for parent interviews (.45–.68; Shaffer et al., 2000). Children's reliability increased with age, which is expected considering their increase in intellectual abilities, greater memory, and improved language comprehension and expression. In contrast, reliabilities based on ratings from interviews with the parents decreased with the child's age, probably because the parents have progressively less contact with their child.

Research on the validity of the DISC has found that discriminations between psychiatric and pediatric groups were good for children with severe diagnoses and severe symptoms but not for children with mild to moderate difficulties (Shaffer et al., 2000).

Discriminations based on interviews with parents were generally more accurate than those based on children (Costello, Edelbrock, & Costello, 1985). Accuracy was also higher for externalizing than internalizing disorders (Friman et al., 2000). In addition, comparisons between psychiatric and pediatric referrals indicated that psychiatric referrals had more symptom scores and more psychiatric diagnoses than pediatric referrals (Costello et al., 1985). The DISC has also been found to identify risk factors for substance abuse (Greenbaum, Prange, Friedman, & Silver, 1991) and to predict behaviors related to conduct and oppositional disorders (Friman et al., 2000). Ratings between DISC and clinician-based diagnosis were moderate to good (.29–.74 for parent and .27–.79 for child; Shaffer et al., 2000) in research settings and followed strict diagnostic guidelines. However, there was very poor agreement between DISC and clinician-based diagnosis when the clinicians performed diagnosis in everyday clinical settings (A. L. Jensen & Weisz, 2002). This lack of agreement may reflect not so much a weakness of the DISC itself but more the fact that there are considerable differences between how diagnosis is achieved in research as opposed to practice contexts. In summary, the DISC has strengths in that it has good reliability and validity among clinical samples involving parent interviews, especially when the problems are related to externalizing disorders. However, the DISC is more problematic when ratings are based on child interviews, particularly among community samples and for internalizing disorders.

Diagnostic Interview for Children and Adolescents

The Renard Diagnostic Interview (Helzer et al., 1981) inspired both the DIS and the DICA (Herjanic & Campbell, 1977; Herjanic & Reich, 1982). The DICA has been through several revisions, which have incorporated the different editions of the *DSM* and elements of the DIS (W. Reich, 2000). Similar to the DIS, the DICA has been designed for administration by lay interviewers. The most recent version (DICA-IV) was published in 1997, aligns with the *DSM-IV*, and is available in child, adolescent, and parent versions (W. Reich, 2000). The DICA can be administered to children between ages 6 and 17 years. The format is semistructured and primarily organized around different themes, such as behavior at home, behavior at school, and interpersonal relationships with peers. Additional content areas are substance abuse and the presence of syndromes such as anxiety disorders, mania, and affective disorders. Elaborate instructions are given for skipping irrelevant items, and total administration time is between 1 and 2 hours. The administration begins with an interview of both the parent and child, which is designed to establish baseline behaviors and to obtain relevant chronological information. The parent is then questioned about the child to determine the possible appropriateness of common *DSM-IV* diagnostic categories. The final step is to administer the “Parent Questionnaire,” which requests additional medical and developmental history and addresses possible diagnoses that have not been covered by previous questioning.

Reliability of the DICA has been quite variable. Test-retest reliability has been quite good, mostly ranging between .76 and .90 (Bartlett, Schleifer, Johnson, & Keller, 1991; Earls, Reich, Jung, & Cloninger, 1988). However, test-retest reliability for child (6 to 12) attention-deficit/hyperactivity disorder was low (.32) and oppositional disorder was

low to adequate (.46; W. Reich, 2000). Reliability has been found to be lowest for questions that were complex, related to time, and for children with the highest level of functional impairment. In contrast, questions with the highest reliability were related to frequency and to externalizing symptoms (Perez, Ascaso, Massons, & Chaparro, 1998). Most cross-informant (parent–child) agreement related to specific symptoms has been disappointingly low (.19 to .54; Herjanic & Reich, 1982). The highest level of agreement was for the oldest children and the lowest was for younger groups (W. Reich, 2000). Whereas parents reported more behavioral symptoms, children were more likely to report subjective complaints.

Validity studies on the DICA indicate that it can accurately make the somewhat gross distinction between middle- to older-age children who were referred to a general psychiatric clinic from those referred to a pediatric clinic (Herjanic & Campbell, 1977). However, there was considerable overlap for children between ages 6 and 8, thus suggesting that a greater possibility of misdiagnosis exists for children in this age range. The DICA was found to be most effective for assessing relationship problems, less effective for academic difficulties, and least effective for assessing school problems, somatic complaints, and neurotic symptoms (Herjanic & Campbell, 1977). In addition, adolescents diagnosed with depression on the DICA also had corresponding elevations on the Beck Depression Inventory (BDI; Martin, Churchard, Kutcher, & Korenblum, 1991). W. Reich (2000) reported that as the genetic similarity of persons diagnosed with Bipolar Disorder decreased, their level of psychopathology on the DISC correspondingly decreased. In summary, the psychometric properties of the DICA have been variable; more studies are needed to substantiate its validity, particularly concurrent validity (R. Rogers, 2001).

RECOMMENDED READING

- Garb, H. N. (2007). Computer-administered interviews and rating scales. *Psychological Assessment*, 19, 4–13.
- Othmer, E., & Othmer, S. C. (2002). *The clinical interview using DSM-IV-TR: Vol. 1. Fundamentals*. Washington, DC: American Psychiatric Press.
- Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York, NY: Guilford Press.
- Sommers-Flanagan, R., & Sommers-Flanagan, J. (2013). *Clinical interviewing* (5th ed.). Hoboken, NJ: Wiley.
- Summerfeldt, L. J., Kloosterman, P. H., & Antony, M. M. (2011). Structured and semistructured diagnostic interviews. In M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 95–140). New York, NY: Guilford Press.

BEHAVIORAL ASSESSMENT

Behavioral assessment is one of a variety of assessment traditions, such as projective testing, neuropsychological assessment, and objective testing. Behavioral assessment distinguishes itself by being a set of specific techniques as well as a way of thinking about behavior disorders and how these disorders can be changed. One of its core assumptions is that behavior can be most effectively understood by focusing on preceding events and resulting consequences. Out of this core assumption has come a surprisingly diverse number of assessment methods, including behavioral interviewing, several strategies of behavioral observation, measurement of relevant cognitions, psychophysiological assessment, and a variety of self-report inventories.

Behavioral assessment can be most clearly defined by contrasting it with traditional assessment. One of the most important comparisons is the emphasis that behavioral assessment places on situational determinants and context of behavior. This emphasis means that behavioral assessment is concerned with a full understanding of the relevant antecedents and consequences of behavior. In contrast, traditional assessment is often focused on the parts of behavior that are the result of enduring, underlying traits. It is this underlying difference in focus of causation that explains most of the other contrasts between the two traditions. An extension of this conceptual difference is that behavioral assessment goes beyond the attempt to understand the contextual or situational features of behavior and, more importantly, concerns itself with ways to change these behaviors. There is a close connection between assessment itself and its implications for treatment. Thus, behavioral assessment can be more direct, utilitarian, and functional than traditional assessment.

The perceived limitations of traditional assessment were a major factor in stimulating the development of behavioral assessment. Specifically, traditional assessment was considered to focus too extensively on abstract, unobservable phenomena that were distant from the actual world of the client. In addition, behaviorists felt that traditional clinical psychology had stagnated because its interventions were not sufficiently powerful and too much emphasis was placed on verbal therapy. The concepts of traditional assessment seemed to exist in an abstract world divorced from the immediate realities and requirements of behavior change. The result of many traditional procedures seemed to be a large quantity of information that had little direct relevance to treatment intervention and outcome. However, this is a stereotyped, somewhat polarized view of traditional assessment; there has been considerable and increasing emphasis on both situational context and treatment implications of information derived from traditional methods of assessment. This stereotyped view highlights differences between the two strategies rather than capturing the complexities and similarities between them.

A further contrast between behavioral and traditional assessment is that behavioral assessment is concerned only with clearly observable aspects in the way a person interacts with his or her environment. A typical behavioral assessment might include specific *measures of behavior* (overt and covert), *antecedents* (internal and external), *conditions surrounding behaviors*, and *consequences*. The clinician then can use this knowledge to specify methods for changing relevant behaviors. Although some behavioral assessors might take selected personality traits into account, these traits would be considered relevant only if they had direct implications for behavior or treatment. For example, certain personality styles interact with the extent and type of depressive cognitions (Alloy et al., 1999), and the existence of a personality disorder typically predicts therapeutic outcome (see Nelson-Gray & Farmer, 1999). This focus on the person and his or her unique situation is quite different from psychodynamic, biochemical, genetic, and normative trait models.

The behavioral approach stresses that different behavior disorders are typically expressed in a variety of modes. These modes might include overt behaviors, cognitions, changes in physiological states, or patterns of verbal expressions. This approach implies that different assessment strategies should be used for each of these modes (Haynes & O'Brien, 2000). An inference based on one mode does not necessarily generalize to another. For example, anxiety for one person may be caused and maintained primarily by the person's cognitions and only minimally by poor social skills. Another person might have few cognitions related to anxiety but be anxious largely because of inadequate social skills. The person with inadequate social skills might be treated most effectively through social skills training and helped only minimally through approaches that alter irrational thoughts (see Breitholtz, Johansson, & Ost, 1999). It should also be noted that altering a person's behavior in one mode is likely to affect other modes, and these effects should be considered.

Whereas the preceding information presents a relatively rigid and stereotyped distinction between traditional and behavioral assessment, most practicing clinicians, including those who identify themselves as behavior therapists, typically combine and adopt techniques from both traditions (Haynes & Heiby, 2004; Hersen, 2005a). Evidence for the high proportion of clinicians combining both traditions is the finding that between 50% and 80% of clinicians who described themselves as being behaviorally oriented reported using structured personality tests, such as the Minnesota Multiphasic Personality Inventory (MMPI; Guevremont & Spiegler, 1990; Watkins, Campbell, & McGregor, 1990). Watkins et al. (1990) even found that about 50% used projective tests and a full 32% used the Rorschach. Thus, behavioral assessment has become increasingly eclectic and is now usually perceived as part of mainstream assessment rather than as a contrasting alternative. Traditional and behavioral approaches have now come to resemble each other in many areas. In particular, behavioral assessment has gone through both a turning inward and a turning outward toward traditional psychometric approaches. It has turned inward in that understanding internal behavioral repertoires and aspects of cognition are seen as essential for a complete understanding of the person (Glass & Merluzzi, 2000; Lodge, Tripp, & Harte, 2000; Nezu, Nezu, Peacock, & Girdwood, 2004). Specific cognitive techniques include having the person think aloud as he or she is involved in a specific situation, sampling thoughts when a beeper goes off, and using a wide variety of self-statement inventories. Behavioral

assessment has turned outward in that it has become increasingly concerned with traditional psychometric considerations. This concern with psychometric evaluation has included assessing the reliability and validity of behavioral observations, self-report inventories, and diagnoses (Haynes, 2006; Hersen, 2005a).

The assumptions and perspectives of behavioral assessment have resulted in an extremely diverse number of approaches and an even wider variety of specific techniques. These approaches and their corresponding techniques can be organized into the areas of behavioral interviewing, behavioral observation, cognitive behavioral assessment, psychophysiological assessment, and self-report inventories. Each of these areas was developed within a wider historical context extending over several decades.

HISTORY AND DEVELOPMENT

Treatment based on behavioral principles has a long history, dating back to the days of Little Albert and his fear of white, furry objects (M. Jones, 1924). However, extensive, well-defined behavioral assessment strategies that were consistent with behavioral therapy were relatively slow to develop. The earliest formal use of behavioral assessment occurred in industrial and organizational settings (Hartshorne & May, 1928; Office of Strategic Services Staff, 1948), but behavioral assessment did not become popular in the clinical context until the mid- to late 1960s. This was probably because of the powerful influence of psychodynamic approaches among clinicians who were taught to “look beneath the surface” to understand the “true” causes of behavior. Perhaps in part as a reaction to this indirect and inferential approach to understanding the person, the earliest forms of behavioral assessment focused almost exclusively on observable behaviors. Although organismic variables, such as cognitions, feelings, and psychophysiological responses, were acknowledged, they were not considered important direct influences on behavior and, as a result, were not stressed in assessment and treatment. Instead, behavioral assessment was consistent with the then-dominant operant conditioning paradigm, focusing on identifying discrete behavioral responses, target behaviors, and reinforcers that could change specific behaviors. Measurement of these areas typically quantified the frequency, rate, and duration of relevant behaviors (Ullman & Krasner, 1965). The result included numerous, highly innovative assessments of overt behaviors. Typically, interventions involved single cases, which was consistent with an idiographic approach.

Early definitions of behavioral assessment were created in part by making contrasts with traditional psychodynamic approaches. Each approach had different aims (identification of problem behaviors versus classification), assumptions (behavior is caused by situations versus enduring traits), and applications (direct observation versus indirect inferences). In particular, Mischel (1968) attacked the very nature of traits by arguing that they were fictions based on distortions of language (a preponderance of static descriptions), the result of consistency of roles and situations, perceptual biases based on needs for predictability, and a bias toward confirmation when traits are (incorrectly) inferred. This attack fueled a lengthy controversy, which was relevant to behavioral assessment in that Mischel’s perspective was used to argue for a focus on situational determinants of behavior. Proponents of behavioral assessment (along with psychiatry

itself) were also dissatisfied with traditional *DSM-II* diagnoses, which had poor reliability and validity and did not seem to relate to the real world of the client or have direct treatment utility.

During the 1970s, there was a much greater emphasis on a wider approach. The typical single case study format gave way to assessment within a much larger context, such as schools, businesses, families, and differing sociocultural frameworks. This assessment approach was based partially on the observation that these larger contexts could have considerable influence on individuals, so that effective individual change often required change in these wider contexts. A refocusing on larger contexts was also motivated by challenges to the strict operant paradigm, which, while effective in controlled situations (hospital ward, Skinner box, prison), had questionable social validity and doubtful long-term clinical impact (Goldfried, 1983; Milne, 1984). Assessment was also widened by arguments to focus on the wider aspects of the person, which meant not only behavior but also feelings, sensations, internal imagery, cognitions, interpersonal relations, and psychophysiological functioning (Lazarus, 1973, 2005). This emphasis on a multimodal or multifaceted approach forced the mainstream of behavioral assessment to accept a number of indirect measures, such as self-reports, ratings by significant others, and cognitions (Cone, 1977, 1978). Relevant publications were the first editions of *Behavioral Assessment: A Practical Handbook* (Bellack & Hersen, 1976), *Handbook of Behavioral Assessment* (Ciminero, Calhoun, & Adams, 1977), and the journals *Behavioral Assessment* and the *Journal of Behavioral Assessment*, both which began in 1979.

The 1980s and 1990s saw a proliferation of publications in the field of behavioral assessment, a dramatic reevaluation of some of its most basic assumptions, and the incorporation of influences from other traditions and disciplines. In particular, psychiatry had similar difficulties with the *DSM-II* as behavioral assessment and began to develop strategies quite similar to those of behavioral assessment. The Problem Oriented Record (Weed, 1968) was introduced into many general hospital and psychiatric settings to improve diagnostic and treatment practices by providing behavior-specific databases, problem lists, treatment plans, and follow-up data. It thereby more effectively tied in the relationship between assessment and treatment, and it more clearly delineated diagnostic issues. Perhaps of greater importance, the *DSM-III-R* and *DSM-IV* were similar to the efforts of behavioral assessment in that each diagnostic category was developed using behavior-specific descriptions. Numerous publications have worked to integrate behavioral assessment with traditional psychiatric diagnosis (First et al., 1992; Follette & Hayes, 1992; Hersen, 2005a, 2005b; Nelson-Gray & Paulson, 1991) in areas such as depression (R. Nelson & Maser, 1988), the diagnosis of childhood disorders (Hersen, 2005b), and personality disorders (Nelson-Gray & Farmer, 1999). The perspectives of psychiatry and behavioral assessment have been further linked by the *Journal of Behavior Therapy and Experimental Psychiatry*.

The development and expansion of behavioral medicine and neuropsychology has also drawn extensively on behavioral assessment strategies in the evaluation of headaches, coronary heart disease, Reynaud's disease, asthma, chronic pain, sleep disturbances, and brain-behavior relationships (Franzen, 2004; Williamson, Veron-Guidry, & Kiper, 1998). In addition, behavioral assessment strategies have focused on complex causal models as well as unstable, transitional behaviors

(S. N. Haynes, 1995; O'Brien, Kaplar, & McGrath, 2004). Thus, not only has behavioral assessment increasingly accepted the contributions of other disciplines and alternative models of conceptualizing behavior, but many of the most honored behavioral techniques have been challenged (Goldfried, 1983). For example, clinical judgment in the context of structured interviews has been generally accepted, diagnostic classification is now considered potentially useful, reliance solely on behavioral observations is perceived in many contexts as inappropriate, and indirect measurement is often seen as essential. In addition, more inferential techniques, such as measuring underlying cognitive structures (schemas) that organize more specific thoughts and behaviors, have become a frequent part of behavioral assessment (Linscott & DiGiuseppe, 1998). This focus on internal aspects of behavior has accompanied a dramatic decrease in the early, time-honored focus on measuring observable frequencies of target behaviors (Glass & Merluzzi, 2000; Guevremont & Spiegler, 1990).

In essence, the 1980s and 1990s witnessed a significant reappraisal and expansion of what is involved in behavioral assessment. In 1989, Birchler summarized his review of behavioral assessment by noting: "Behavioral assessment as we may have known it in the recent past is in a rapidly changing process of (choose one): disarray, revision, broad expansion, advancement, confusion, and/or extinction" (p. 385). There has certainly been a significant blurring and cross-fertilization between behavioral assessment and other forms of assessment (Haynes & O'Brien, 2000), which is reflected in part in the fact that the *Behavioral Assessment* journal and the *Journal of Behavioral Assessment* have changed their names and content to include wider aspects of psychopathology and more traditional assessment tools (e.g., MMPI-2; Millon Clinical Multiaxial Inventory [MCMI-IV]). Current directions include analog and virtual reality assessment, advances in psychophysiological assessment, utilization of innovative technology, use of ambulatory sensors, assessment of clients in their natural environments, and greater applications for special populations (Haynes & Yoshioka, 2007; Hersen, 2005a, 2005b; Piasecki, Hufford, Solhan, & Trull, 2007). Future directions that have been highlighted by Ollendick, Alvarez, and Greene (2004) are the need to focus more on developmental factors, incremental validity when using multiple assessment methods, and the inclusion of culturally sensitive approaches.

ISSUES RELATED TO RELIABILITY AND VALIDITY

Traditional psychometric considerations for behavioral assessment are difficult to summarize because of the wide diversity of techniques and the differences in assumptions regarding the focus, nature, and causes of behavior. Whereas traditional assessment stresses the relative stability of various characteristics, behavioral assessment assumes variability based largely on environmental and contextual factors. A finding such as low test-retest reliability is more likely to be interpreted in the behavioral context because of true variance resulting from environmental conditions rather than error in the data collection procedure. Furthermore, behavioral assessment stresses the importance of individually tailored approaches emphasizing the client's idiosyncrasies. In this context, normative comparisons are frequently seen as both irrelevant and inappropriate. Despite these issues, many from within the area of

behavioral assessment have successfully argued for evaluating behavioral assessment techniques with traditional psychometric approaches (Cone, 1998; Haynes, 2006). For example, interobserver agreement for behavioral observations is essential before the data gathered from this approach can be trusted. Interobserver agreement is typically evaluated by calculating the percentage of interrater agreement. Likewise, data derived from self-reports in areas such as assertiveness and fear need to demonstrate that the findings can be generalized to other situations, such as role plays, simulations, and, especially, daily life.

The earliest forms of behavioral assessment relied primarily on behavioral observation and assumed that the direct observation of specific behaviors was sufficiently clear, reliable, and accurate. The emphasis was primarily on determining a functional analysis between behavior and its antecedents and consequences. In an activity such as pressing a bar for reinforcement, the behavior could be easily recorded by an electronic detector; therefore, the reliability of the measure could be considered to be quite high. However, with behaviors that are more difficult to define, the reliability of measurement, especially when based on behavioral observation, cannot be assumed. For example, fingernail-biting might be defined merely by the person touching his or her face, or it may involve touching the mouth, actually chewing the nail, or removing part of or the entire nail. The issue of precise definition and accurate measurement of the behavior becomes even more problematic when dealing with internal cognitions, for which the clinician is completely dependent on self-report rather than direct observation.

The level of reliability across different observational strategies has varied. In general, material derived from behavioral observation during behavioral assessment can be influenced by observer expectations, as has been found by experimental research (H. M. Cooper & Rosenthal, 1980; Orne, 1962; R. Rosenthal, 1966). Consistent with this fact is the finding that interrater agreement has been quite variable for areas such as overt difficulties and underlying mechanisms (Persons, Mooney, & Padesky, 1995). In situations such as natural observation, in which observer bias, outside factors (such as interference from nontarget persons), and a lack of clear definitions are likely to create variability in observer ratings, reliability can be expected to be relatively low. Further sources of observer error include halo effects, primacy effects, failure to score a behavior that has occurred, rating toward the center of the scale, and leniency or generosity of scoring. When bias is reduced by using highly structured and standardized procedures, reliability increases. Thus, a procedure such as systematic sampling, in which clear strategies are used to determine when and how the behavior will be measured, has generally been more reliable and accurate than naturalistic observation (Cunningham & Thorp, 1981). Although reliability has been found to increase in controlled situations where the observers know that they themselves are being evaluated for accuracy (Romanczyk, Kent, Diamant, & O'Leary, 1973), this outside monitoring of observers rarely occurs in clinical situations. Thus, it cannot be assumed that the reliability found in clinical situations is as high as for controlled studies in which evaluators are themselves being evaluated. General guidelines for increasing reliability in clinical situations include having two observers compare their results, providing careful instructions when a client is asked to monitor his or her own behavior, specifying target behaviors, clearly wording items on self-reports, taking care in the construction of instruments, and thoroughly training observers such as parents or teachers. Reliability

of ratings is also likely to increase by clinicians paying closer attention to contextual variables (J. G. Beck, 1994; Haynes, 2006; Haynes & O'Brien, 2000).

During the 1960s and 1970s, the validity of various assessment procedures depended primarily on informal content validity. Questionnaires and observational strategies were based on rational considerations regarding what was to be studied and how these measurements were to be made. Few efforts were made to develop empirically derived categories. For example, the assessment of depression might have been based on knowledge about the typical thoughts depressed people seem to have as well as additional variables that seem important regarding social supports and typical antecedent events. The various areas of observation were selected mostly based on what rationally seemed to be the most critical considerations. Since the early 1980s, increased work has gone into assessing the validity of various methods of behavioral assessment. In general, few validity studies have been performed on behavioral interviews and naturalistic observations, whereas much more has been done on behavioral questionnaires. Most validity studies have been conducted by using relevant outside criteria. Many of the same issues have come up with criterion validity for behavioral assessment as for traditional assessment, including difficulty generalizing to different populations, settings, and methods of administration.

The early behavioral self-report questionnaires relied on content and face validity. Because these questionnaires represented new techniques with a different underlying philosophy, it was believed that they did not have to be judged using the same criteria as the older and more traditional psychometric tests. They were considered to be direct reports of client behaviors, and thus little psychometric validity was reported. R. M. Kaplan and Saccuzzo (1993) criticized this initial focus on content and face validity by stating that behavioral self-reports may be "repeating history and reinventing the wheel" (p. 493). They further pointed out that the "early paper-and-pencil structured personality tests which were finally abandoned in the 1930s are indeed difficult to distinguish from many present-day (behavioral) self-report procedures" (p. 494). Many early behavioral self-report questionnaires could have been best described as "idiosyncratic clinical tools" rather than psychometrically sound tests. The problems of response bias, questionable reliability and validity, lack of norms, and assumed client truthfulness need to be addressed for any standardized instrument, including behavioral procedures.

As a result of these criticisms, much greater attention was focused on behavioral techniques, especially self-report inventories. However, the success of these efforts has been quite variable. For example, the Rathus Assertiveness Schedule (RAS; Rathus, 1973) has been subjected to traditional psychometric procedures and illustrates the difficulties encountered in this and other similar behavioral inventories. Whereas Heimberg, Harrison, Goldberg, Desmarais, and Blue (1979) did not find a very high correspondence between scores on the RAS and observational reports of role plays in an inmate population, the RAS did relate to nonassertiveness in a group of dental students (Rathus, 1972). However, a difficulty with relating assertiveness in role-play situations, which most of the preceding studies used, is that assertiveness in role plays may not relate to assertiveness in naturalistic situations (Bellack, Hersen, & Turner, 1979). Perhaps when subjects are asked to role-play, they can alter their daily level of assertiveness to "act the part" correctly (Higgins, Alonso, & Pendleton, 1979).

The RAS similarly has poor criterion validity based on instructor evaluations of observed assertive behavior and grades in a communication course (Tucker, Weaver, Duran, & Redden, 1983). Thus, even though the RAS is a frequently used device in both research and clinical settings, the meaning of the scores might be difficult to evaluate. Other behavioral self-report questionnaires have experienced similar problems.

ASSETS AND LIMITATIONS

Probably the greatest advantage of behavioral assessment is that its practitioners have continually paid attention to its relevance toward treatment. Measurement of problem behaviors is usually directly tied to how these behaviors can be changed. Furthermore, relevant behaviors are given an empirical functional analysis, which enables clinicians to make baseline measurements of behavior and to assess the antecedents and consequences of these behaviors. An initial functional analysis can then allow clinicians to evaluate whether change has actually occurred during or after treatment. Although many techniques have not been through rigorous traditional validity studies, the emphasis on treatment utility has proven to be attractive to many practitioners. Thus, behavioral assessment is particularly useful for persons using a hypothesis-testing approach and for those who wish to have clear accountability that behavioral change has actually taken place. In some situations, however, behavioral assessment can be tied too closely to treatment, particularly in legal assessments or other situations in which assessment and therapy should be separate.

A further asset is that behavioral assessment offers a wide range of possible techniques for use in extremely varied contexts. These strategies include self-reports, naturalistic observation, physiological monitoring, ambulatory sensors, structured observation, and self-monitoring. Variations in techniques are consistent with the view that a complete understanding of the person requires multiple modes of assessment. The different assessment modes might involve relevant aspects of person–situation interaction, physiological changes, cognitions, interpersonal relationships, overt behaviors, feelings, imagery, and aspects of the person's larger social system. Many behavioral assessment models organize their approach around stimulus, organism, response, and contingencies (Goldfried, 1982). Other approaches rely on Lazarus BASIC-ID (Lazarus, 2005) or on Kanfer and Saslow's (1969) functional analysis of behavioral excesses and deficits. These approaches place the person in a much wider context than traditional assessment procedures.

Behavioral assessment is particularly appropriate when a presenting problem is likely determined primarily by environmental factors. For example, in most cases, a clear, functional relationship (environmental interaction) can be established for phobias, marital difficulties, acting out, temper tantrums, and inappropriate classroom behavior. If these behaviors are frequent in occurrence (i.e., smoking, classroom acting out), it is fairly easy to develop a baseline and monitor change. However, quite unique behavior that occurs infrequently (e.g., relapse into substance abuse, bringing firearms to school) may be much more difficult to measure and monitor (J. R. Nelson, Roberts, Rutherford, Mathur, & Aaroe, 1999). In addition, behavioral assessment is somewhat less relevant when environmental factors account for a smaller portion of the variance.

In such cases, organic factors may be more important than environmental ones, such as in chronic schizophrenia, certain types of headaches, and head injuries. Although behavioral assessment and intervention can still be effective for such problems, greater difficulties are involved because the environment is relatively less important.

A previously described but extremely important drawback of many behavioral assessment strategies is that they have poor or, at least, untested psychometric properties. Often attempts to establish reliability and validity have been disappointing. In addition, the accuracy of behavioral observation and interviewing can be distorted because of observer bias, halo effects, primacy effects, low interobserver agreement, confirmatory bias, and so forth.

Although cognitive behavioral assessment has gained increased importance over the past 30 years, in many ways it is contrary to the original spirit of behavioral assessment's emphasis on direct observation. Cognitive assessment is necessarily unobservable and relies on client self-report. Difficulties might include differences in meaning between the client and the clinician, response biases, assumed honesty of reporting, and assumptions about the equivalence of internal dialogues and their verbal descriptions.

A final limitation of behavioral assessment is that it often requires extensive resources in terms of time, personnel, and equipment. This is particularly true for psychophysiological and observational methods. Surveys indicate that while clinicians do use direct observation, they do so only on an occasional basis (Elliot, Miltenberger, Kaster-Bundgaard, & Lumley, 1996; Guevremont & Spiegler, 1990). In contrast, they are much more likely to use interviews with the client, interviews with significant others, self-monitoring, and behavioral rating scales. As a result, behavioral assessment is often limited to interviews and questionnaires (Guevremont & Spiegler, 1990; Sayers & Tomcho, 2006). An additional drawback is that many behavioral instruments have not been designed to deal with problems frequently encountered in clinical practice, such as dissociative disorders, paranoia, and hypochondriasis.

STRATEGIES OF BEHAVIORAL ASSESSMENT

Behavioral assessment has given rise to numerous and highly varied techniques, many of which are described in *Comprehensive Handbook of Psychological Assessment (Volume 3): Behavioral Assessment* (Haynes & Heiby, 2004); *Clinician's Handbook of Adult Behavioral Assessment* (Hersen, 2005a); *Clinician's Handbook of Child Behavioral Assessment* (Hersen, 2005b); and *Dictionary of Behavioral Assessment Techniques* (Hersen & Bellack, 2002). For example, an estimated 300 questionnaires can be conceptualized as behaviorally oriented self-report instruments (Hersen & Bellack, 2002). Despite this diversity, behavioral assessment strategies can be organized into the general categories of behavioral interviewing, behavioral observation, cognitive behavioral assessment, psychophysiological assessment, and self-report inventories. Each of these approaches varies in the degree to which it emphasizes direct versus indirect measures of the person as well as the extent to which it relies on inference. For example, cognitive assessment is more indirect than behavioral observation and relies much more on inferences regarding the degree to which cognitions affect and interact with overt behavior. However, all of these techniques stress developing a functional

analysis of behavior through understanding person–environment interaction. They also emphasize that each aspect of assessment is directly relevant to treatment planning and evaluation.

Behavioral Interviewing

Behaviorally oriented interviews generally focus on describing and understanding the relationships between antecedents, behaviors, and consequences (ABC). In addition, a baseline or pretreatment measure of behavior is developed through a systematic consideration of the frequency, intensity, and duration of relevant behaviors. Behaviors might also be evaluated with a description of specific behavioral excesses and deficits (Kanfer & Saslow, 1969). Any goal must be capable of being measured and tested in an objective and reliable way, and the client should agree on its relevance (Gresham, 1984). Although the behavioral approach might seem long and involved, the process is simplified by considering only areas that are relevant for treatment.

Despite this emphasis on treatment utility, it is essential to place each aspect of the information derived from a behavioral interview into a broad context. A basic description of a target behavior is too simplistic because it does not take into account an interactionist model. For example, a phobia is likely to create difficulties in the client's relationships, which could undermine the person's sense of competence. The person might then react by becoming highly dependent on a primary relationship, reinforcing the sense of helplessness. The helplessness might then reinforce a fear of not being able to cope, which can then interact with and quite possibly exacerbate the phobia. Thus, a complete interview would evaluate not only the existence of and nature of the phobia but also the effect of the phobia on relationships, work effectiveness, and self-statements. Whereas the earlier behavioral interviews of the 1960s and 1970s often had a narrow focus, current models of behavioral assessment emphasize taking this wider context into consideration (Nezu et al., 2004).

The general purpose of the behavioral interview is multifaceted. It might help identify relevant target behaviors or select additional behavioral assessment procedures. It also provides an opportunity to obtain informed consent, obtain a history of the problem, identify potential causal factors related to the presenting problem, develop a functional analysis of the problem behavior, increase client motivation, design intervention programs, and evaluate the effectiveness of previously attempted interventions.

The initial phase of a behavioral interview needs to include many of the elements relevant for traditional interviews. A sufficient degree of rapport needs to be established, a statement needs to be developed of the general and specific purposes of the interview, and a review should be made of the client's relevant history. However, history tends to be deemphasized in favor of current behaviors because the main cause of client behavior is considered situational rather than historical. Common clinician approaches involve reflective comments, probing, conveying understanding, and expressed empathy. Open-ended questions can be followed up with more direct questioning. However, the extensive use of nondirective techniques is inappropriate; the clinician must set a clear direction and have the client answer direct questions relevant to a behaviorally oriented approach.

Sometimes clients provide excellent descriptions of their problems and can specify relevant antecedent and consequent conditions. Other clients experience difficulty describing the events surrounding the decision to seek treatment, elaborating on their feelings, providing information about how other people might be perceiving their problem, or even stating who referred them. Because a careful behavioral analysis requires a complete description of problem behaviors, the client and therapist must work to establish the extent of the difficulty, where it occurs, when it occurs, and its effects on relationships. Sometimes it is helpful to have the client keep a diary of relevant events and observations. New technologies using electronic devices are available to assist with ongoing self-monitoring (Piasecki et al., 2007). Often clients describe and define their difficulties by relying extensively on general trait descriptions rather than on more behaviorally oriented ones. A behavioral interviewer, then, needs to work with the client to develop specific and easily observable descriptions. For example, if a client says he or she is a “depressed type of person,” this might translate into specific types of behaviors (slow movement, spending too much time in bed, avoiding people, being nonassertive), cognitions (that he or she is no good, a failure), or feelings (hopelessness, apathy). The assumption of an underlying permanent trait (illness) needs to be reframed as a group of specific behaviors that are potentially changeable. This reframing process, in itself, is likely to be beneficial to clients, as they will be better able to see specific things they can do to change how they feel. Speaking in concrete behavioral terms rather than abstractions is also likely to increase mutual understanding between client and therapist.

A wide-based behavioral assessment should describe not only the specific presenting problem but also the manner in which the problem has generalized into other areas. In particular, the assessment might involve information about the larger social system. Often the client’s school, work, or family situation can be incorporated into the assessment and treatment program to ensure both immediate and long-term success. In contrast, if a narrow approach to change is taken, the client may attempt to express his or her newly acquired behavior in contexts that are not supportive of it. As a result, previous problem behavior might once again emerge to the exclusion of newer, more adaptive behavior. This might be true if the client developed new, effective behaviors that were learned only in the narrow context of the practitioner’s office.

An interview should end by providing the client with a summary of the information obtained, an explanation of additional information that is required, and an estimate of the likely success of treatment (Sayers & Tomcho, 2006). If further information is required, the clinician and client should agree on what is needed and how to obtain it. This need for additional information might involve instructions for keeping an effective diary, requests for observations from other people, or techniques for self-monitoring of different behaviors. If the interview is a prelude to therapy, additional information should be given about possible strategies for intervention, the length of treatment, possible financial and emotional costs, and assurances that the client will have input into all decisions.

Because most interviews tend to be somewhat informal and haphazard, they frequently provide information with low reliability and validity. For example, T. Wilson and Evans (1983) found a low level of reliability among clinicians trying to specify

appropriate target behaviors. Some authors urge that behavioral interviews be structured and standardized. Kratochwill (1985) suggested that interviews be planned around a four-stage problem-solving process. The first stage is *problem identification*, during which the problem is specified and explored and procedures are established to measure current performance and desired target behaviors. The vague and generalized descriptions that clients typically come in with are developed into specific behavioral descriptions. Next, a *problem analysis* is performed by assessing the client's resources and by noting the relevant environmental conditions influencing behavior and the context in which the behavior excesses or deficits occur. An interview also needs to establish how a *plan might be implemented*, which would include ongoing procedures for collecting data relevant to the progress of the treatment. Finally, strategies for *treatment evaluation* should be specified by considering the pre- and posttreatment measures to determine whether the intervention was successful.

Witt and Elliott (1983) provided this somewhat similar outline of expected accomplishments for any behavioral interview:

1. Initially, provide the client with an overview of what needs to be accomplished and why a clear and detailed specification of the problem behavior is important.
2. Identify the target behavior(s) and articulate them in precise behavioral terms.
3. Identify the problem frequency, duration, and intensity ("How many times has it occurred today," "How long has it been going on," etc.).
4. Identify conditions in which the problem occurs in terms of its antecedents, behaviors, and consequences.
5. Identify the desired level of performance and consider an estimate of how realistic this is and possible deadlines.
6. Identify the client's strengths.
7. Identify the procedures for measuring relevant behaviors: What will be recorded, who will record it, how will it be recorded, when and where will it be recorded?
8. Identify how the effectiveness of the program will be evaluated.
9. After completing discussion of the preceding areas, summarize it to ensure that the client understands and agrees.

This outline should not be followed rigidly but should be used as a general guideline. However, each behavioral assessment should have accomplished all nine areas before completion.

Behavioral Observation

In some cases, the behavioral interview is itself sufficient to obtain an adequate assessment. However, some form of actual behavioral observation is often required before, during, and/or after treatment. The particular method for observing behavior is usually decided on during the initial interview. Whereas the interview is directed primarily toward obtaining verbal information from the client, behavioral observation is used to decide on and actually carry out specific strategies and techniques of measuring the relevant areas of behavior discussed during the interview (see Suen & Rzasa, 2004).

In some cases, such as assessing those with developmental disabilities, resistant clients, or very young children, behavioral observation may become one of the most important means of assessment. These observations might be made by the professional who is actually conducting the treatment or by someone else who is more involved in the client's life, such as a teacher, parent, or spouse, or by self-monitoring by the client. The most frequent approaches are narrative recording, interval recording, event recording, and ratings recording.

The first behavioral observation task is to select relevant target behaviors, which can vary from a single response set to a larger interactive unit. The target behavior should either involve the problem behavior itself or relate to it in a meaningful way. Decisions must be made regarding the number of behaviors to record and the relative complexity of the recording method. Both the recording method and the target behavior need to be manageable and not overly complex. The target behavior can best be situated by beginning with a narrative description of the client's difficulty; it can be further specified by considering the antecedents and consequences related to the problem behavior.

All behaviors to be measured must have objective, complete definitions (operationalization) that allow clear observations of the measures of the behavior. In particular, the definition should avoid abstract and highly inferential terms, such as *apathy* or *sadness*, and instead translate such terms into specific behaviors. Any description of the target behavior should involve an easy-to-read dictionary-type definition, an elaboration of the behavior, and specifications regarding precisely when the behavior occurs, as well as descriptions of borderline examples and clear nonexamples. In measuring behavioral frequencies, the observer must clearly define when the discrete behavior begins and ends. Doing this might be easy for measuring the number of cigarettes a person smokes or the number of times a child bangs his or her head, but it is more difficult when measuring less clearly defined behaviors, such as the number of aggressive acts a person makes or frequency of nonassertive behaviors. Recordings should also measure the duration of behaviors and their intensity. For example, how hard a child bangs his or her head and the total time engaged in the activity have implications for the urgency and strength of the treatment approach.

The different devices used to make recordings might include various combinations of golf counters, stopwatches, pencil-and-paper forms, or electromechanical devices, such as an event recorder with buttons that can be pressed when various categories of behaviors occur. Smartphones are becoming more common, as most have built-in capability for recording.

The settings of behavioral observation can range from those that are natural to those that are highly structured. Natural, or in vivo, settings might include the home, classroom, business, or playground. Observations made from these types of settings are likely to be directly relevant to and reflective of the client's life. Natural settings are most effective when assessing high-frequency behaviors and/or more global behaviors, such as attentional deficits, social withdrawal, or depressive behaviors. They are also useful when measuring the amount of change the client has made following intervention. However, natural settings present difficulties because of the extensive time required to make observations. Furthermore, natural settings are problematic when trying to measure infrequently occurring behaviors (aggression, nonassertiveness) or behaviors that typically occur in the absence of others (fire-setting, suicide). To counter the difficulties

inherent in naturalistic observation, practitioners may wish to create analog situations, such as role plays or work simulations, that elicit specific types of behaviors. Such environments are especially important for infrequent behaviors. However, inferences need to be derived cautiously from observations in these structured or analog situations, as they may not generalize well to the client's actual life.

When clinicians are concerned that having an observer not readily known to the client or the client's environment might contaminate the results, they may wish to train persons who are already a part of the client's natural setting, such as parents, teachers, or spouses. This method might help prevent subjects from changing their behaviors simply because they are aware that they are being observed ("reactivity"). These more natural observers can be much less obtrusive than an outside professional. The training of observers needs to include a clear rationale for measuring the behavior with emphasis on making accurate and objective recordings. Observers should memorize the recording code, practice making the recordings, and receive feedback about the relative accuracy of their recordings. Precautions should be taken to avoid observer error, such as through observer bias, leniency, lapses in concentration, and discussion of data with other observers or with the client. Sometimes reliability might be checked by comparing the degree of agreement between different observers rating the same behaviors. Trained observers should be used cautiously because widely varying levels of interobserver agreement have been noted (Margolin, Hattem, John, & Yost, 1985).

A system of coding behaviors usually needs to be developed so that recordings are abbreviated and simplified. If too many codes are used, it is difficult for recorders to recall them, especially if behaviors occur in rapid succession. Both the type of recording method (narrative recording, event recording, etc.) and the coding system depend largely on the goals of assessment. A coding system that is clear, simple, and closely connected to the presenting problem is likely to be both useful and reliable. Important considerations in selecting a recording and coding system are the number of times the behavior needs to be observed, the length of observation periods, when to make the recording, the type of recording to be made, and the target behaviors to be recorded. The following sections describe the most frequently used recording systems, along with examples of different methods of coding.

Narrative Recording

Narrative recording requires that the observer simply make note of behaviors of interest. There is little quantification, and the observations can vary in the degree of inferences made. For example, an observer may stick closely to direct descriptions of behavior, such as noting that someone frequently laughs and smiles at his or her friends, or may infer from these behaviors that the client has positive peer relations. The primary value of narrative recordings is that they may help define future, more specific areas that can then be measured in a more quantitative, systematic manner. Thus, narrative recording is usually a precursor to alternative forms of measurement. It has the advantages of potentially discovering relevant behaviors; it can elaborate on these behaviors; it requires little, if any, equipment; and numerous hypotheses can be generated from the narrative descriptions. Limitations are that it does not enable the observer to quantify the observations, it may have questionable validity, and the usefulness of the observations depends largely on the individual skill of the observer.

Interval Recording

A clinician may choose to record whether selected behaviors occur within predetermined intervals. As a result, this technique is also referred to as *time sampling*, *interval sampling*, or *interval time sampling*. Usually the intervals vary from 5 to 30 seconds and may be based either on set schedules for each observation period (e.g., every 5 minutes) or may be selected randomly. Interval recording is most appropriately used for measurement of overt behaviors with moderate frequencies (e.g., once every 5–20 seconds) and when these behaviors do not have any clear beginning or end. Examples might include behaviors such as walking, listening, playing, reading, or looking up and down.

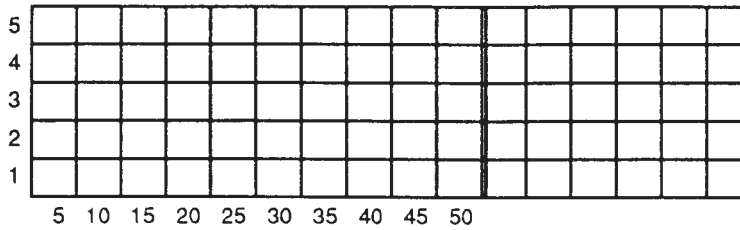
When developing a strategy for interval recording, clinicians must decide on the length of time between each observation, the method of recording, and the length of the observation period. Deciding on which strategy to use depends largely on the type of behavior. For example, different types of verbal interaction vary in length; for that reason, the observation periods must be adjusted. Some strategies might require the observer to alternate between recording (e.g., for 10 seconds), then observing (e.g., for 20 seconds), and then going back to recording the observation just made. Cues regarding the beginning and end of each behavior must be specified. The target behaviors for observation are derived from information based on such sources as the initial interview, self-report inventories, narrative observations, and especially descriptions of the presenting problem. The focus of observation may also vary between different people, such as husband, wife, teacher, child, or client. Sometimes clinicians or researchers arrange to have an outside person observe the same client behaviors. The interrater reliability of the observations can then be established by calculating the percentage of agreement between the two raters (see Suen & Rzasa, 2004). A representative interval recording chart, with instructions on how to develop such a chart, is provided in Figure 4.1.

Interval recording is time efficient and highly focused on specific behaviors, and it has the potential to measure almost any behavior. Interval recording is not designed to assess the quality of the target behaviors, however, and may overlook other important behaviors.

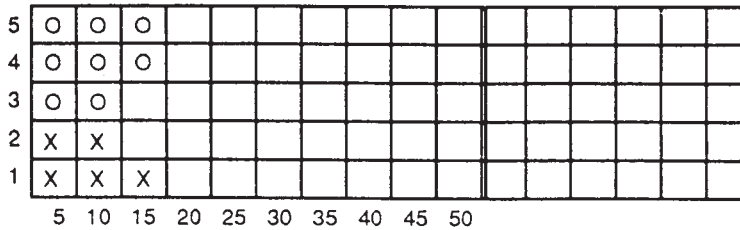
Event Recording

Whereas interval recording depends on measurements defined by units of time that are imposed on target behaviors, event recording depends on the occurrence of the behavior itself. The observer must wait for the target behavior to occur and then record relevant details of the behavior. Examples of behaviors most appropriate for event recording are aggressive actions, greetings, or use of verbal expressions such as assertion or profanity.

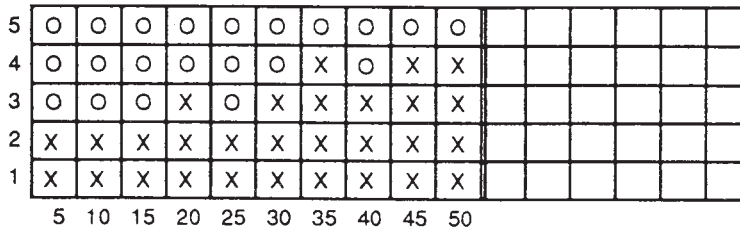
The basic design of event recording systems is to note the behavior's frequency, duration, and intensity and to record the behavior on such devices as a checklist, golf counter, personal digital assistant (PDA), and now even a smartphone. Although the main emphasis is on quantifying the frequency of responding, its duration also can be measured with a timer. The intensity of the behavior can be noted by simply specifying whether it was slight, moderate, or strong. A representative example of an event-recording chart is included in Figure 4.2.



a. Graph paper with series of columns, each five blocks high. Double heavy line marks off 10 columns, for a 50-minute period.



b. Chart after 13 minutes of monitoring pupil's behavior. First two columns are completed and the third is partially completed. If the pupil behaves appropriately during the next (14th) minute, the observer will mark an "X" in the third column just above the other "X." If the pupil misbehaves, the observer will mark an "O" in that column just under the other two "Os."



c. Chart after observer has completed the 50-minute period.

Figure 4.1 Example of interval recording

		Intervals in minutes						
Behavior	Totals	Person Observed	5	10	15	20	25	30
Getting out of seat	27	Subject	□	□	□	□	□	□
	8	Comparison	□	□	□	□	□	□
Requesting help	5	Subject	□	□	□	□	□	□
	11	Comparison	□	□	□	□	□	□

Figure 4.2 Example of event recording within 5-minute intervals

Event recording is especially good for recording behaviors having low frequencies, measuring changes in behaviors over time, and studying many different types of behavior. However, event recording is relatively poor at measuring behaviors that do not have clear beginnings and endings, and it presents difficulties in keeping the attention of observers for behaviors of long durations. Because event recording does not provide information regarding sequences of behaviors, it is difficult to make inferences about how and why behaviors occur.

Ratings Recording

Rather than recording direct observations of behaviors, clinicians may wish to obtain general impressions of relevant dimensions of behaviors and have these impressions rated on a checklist or scale. Such measures tend to be more global and may involve more abstract terms, such as the client's level of cooperativeness or ability to maintain self-care. Typically, ratings recordings are made after a period of observation. A typical format might request the evaluator to rate, on a scale from 1 to 5 or 1 to 7, the client's frequency of temper tantrums, quality of peer relations, or exhibited conscientious behaviors. For example, the Motivation Assessment Scale (MAS; Durand & Crimmins, 1992) is a 16-item questionnaire that evaluates the functional significance of behavior related to the dimensions of sensory, escape/avoidance, social attention, and tangible rewards. Although the MAS is a frequently used instrument, the level of interrater agreement has been found to be quite variable, as has its internal consistency and factor structure (see Kearney, Cook, Chapman, & Bensaheb, 2006).

Ratings recordings can potentially be used for a wide variety of behaviors. Other advantages include the ability for the data to be subjected to statistical analysis; the ratings' ability to be made for either individuals or groups; and because of the time efficiency of ratings recordings, their cost-effectiveness. Disadvantages include possibly low interrater agreement because of the subjectivity of the ratings; little information regarding antecedent and consequent events; and possibly inaccurate ratings, especially if much time elapses between making the observations and making the ratings.

Cognitive Behavioral Assessment

Over the past 30 years, considerable research has been conducted on understanding the cognitive processes underlying behavior disorders. Relevant areas include the self-statements associated with different disorders, the underlying structure or cognitive organization related to these disorders, differences between cognitive distortions in pathological versus normal behavior, and cognitive alterations that occur during therapy. This research has considerably influenced and altered the nature of behavioral assessment. In particular, researchers have developed specific techniques for assessing cognitive processes, such as having the person think aloud, listing different thoughts, thought sampling at various intervals, and a wide variety of self-statement inventories.

This internal perspective is quite different from the early emphasis of behavioral assessment, which focused almost exclusively on observable overt behavior. This transition has come about because of persuasive evidence for the relationship between behavior and cognitions (Alloy et al., 1999; Bandura, 1986; Garratt, Ingram, Rand, & Sawalani, 2007; Haaga, Dyck, & Ernst, 1991; Kendall & Treadwell, 2007). Cognitive

processes not only change during the course of effective therapy but may be causally related to both the development and the maintenance of different types of disorders (Alloy et al., 1999; Breitholtz et al., 1999; Brewin, 1996; Garratt et al., 2007). Some approaches assume that altering cognitions can be sufficiently powerful to change behaviors. However, there are also a number of significant limitations with cognitive behavioral assessment. All material is necessarily derived from the client's self-report of his or her internal processes and, as such, may be subject to a number of distortions. Clients can usually recall and describe the results of their cognitive processes, but they have much greater difficulty describing how they arrived at these conclusions. The actual processes may need to be inferred based on complicated analyses of the results derived from intricate assessment strategies. In addition, remembering events seems to be a reconstructive process in which each successive recall can be altered based on the person's needs, biases, and expectations (Henry et al., 1994; Lindsay & Read, 1995; Loftus, 1993). These inherent difficulties have led some traditional behaviorists to question the theoretical and practical appropriateness of cognitive assessment.

A relevant finding is that the popular belief in the power of positive thinking is simplistic because it is not a very good predictor of adjustment. What seems more important is the absence of negative statements or what Kendall and Hollon (1981) referred to as "the power of nonnegative thinking." Furthermore, the effect of negative self-talk is greater than the ability of positive thinking to counter negative internal dialogue. As might be expected, gains in therapy have been associated with reductions in negative self-statements (Garratt et al., 2007; Kendall & Treadwell, 2007). Another issue is that relevant cognitions such as self-efficacy vary across cultures. For example, many non-Western cultures have been found to have lower levels of self-efficacy; this is particularly true among persons from more collectivist cultures (Klassen, 2004). Despite this finding, levels of self-efficacy were actually more predictive of relevant behaviors among these cultures. Clinicians conducting cognitive and other forms of assessments need to take these contextual variables into consideration.

The two major strategies of cognitive assessment are through various self-report inventories and techniques of recording cognitions. Each of these general strategies has strengths and weaknesses and is appropriate in different situations for different types of clients.

Cognitive Self-Report Inventories

There has been a tremendous expansion in the number and frequency of use of cognitive self-report inventories. Guevremont and Spiegler (1990) noted that they were used nearly as frequently as behavioral interviewing and twice as often as direct observation, though this research is now dated. They have the general advantages of having strong face validity and are both easy and inexpensive to administer. However, their psychometric properties vary greatly, and many instruments in frequent use are quite poor in this regard. Typically, they involve between 20 and 100 items, with respondents asked to indicate their degree of endorsement of each item on a Likert-type scale. Many instruments have been tailored toward specific domains, such as depression, fears and anxieties, self-efficacy, imagery, social skills (especially assertiveness), eating disorders, and marital problems. The main domains for cognitive self-report inventories and the

Table 4.1 Cognitive Self-Report Measures

Domain	Instruments
Depression	Dysfunctional Attitudes Scale Cognitive Bias Questionnaire (child and adult versions) Automatic Thoughts Questionnaire Beck Depression Inventory–II Attributional Styles Questionnaire
Fears and Anxieties	Social Avoidance and Distress Scale Fear of Negative Evaluation Scale Social Interaction Self-statement Test Irrational Beliefs Test Rational Behavior Inventory Fear Survey Schedule
Eating Disorders	Eating Attitudes Test Bulimia Test–Revised Cognitive Error Questionnaire (modified for eating disorders)
Social Skills	Rathus Assertiveness Inventory Wolpe-Lazarus Assertion Inventory Gambrill Assertion Inventory Bakker-Assertiveness Schedule Conflict Resolution Inventory Survey of Heterosexual Interactions Stanford Shyness Scale
Marital Relationships	Relationship Attribution Measure Relationship Belief Inventory Dyadic Attribution Inventory Marital Attitude Survey Specific Relationship Standards

most frequently used instruments in these domains are summarized in Table 4.1. It is beyond the scope of this chapter to review them, but useful information can be obtained in Bellack and Hersen (1998) and Hersen and Bellack (2002).

Theories of the cognitive processes of *depression* suggest that it is maintained by characteristic and repetitive thoughts that are self-perpetuating (Alloy et al., 1999; Garratt et al., 2007). A. T. Beck (1967) originally listed the cognitions associated with depression as involving *arbitrary inference* (making inferences without substantiating evidence), *selective abstraction* (making a broad judgment based on a minor aspect of an event), *overgeneralization* (extrapolating in an unjustified fashion from a minor event), and *magnification/minimization* (overemphasizing negative events; minimizing positive ones). Although these processes seem to be related to depression, a simple cause-effect model between depression and specific cognitions does not appear to be warranted, and further clarification is required (Greenberg, 2008). Some of the most frequently used inventories to assess depressogenic cognitions are the Dysfunctional Attitudes Scale (A. Weissman & Beck, 1978), the Cognitive Bias Questionnaire (Hammen, 1978; Hammen & Krantz, 1976), the Automatic Thoughts Questionnaire

(Hollon & Kendall, 1980; Ingram et al., 1995), and the Beck Depression Inventory (BDI-II; A. T. Beck et al., 1996). More extensive coverage of the BDI-II can be found in Chapter 13. In addition, the Attributional Styles Questionnaire (Seligman, Abramson, Semmel, & von Baeyer, 1979) is sometimes used to better understand the manner in which a client construes the causes for various behaviors, particularly those related to depression (e.g., learned helplessness).

A wide number of measures have been developed related to a person's *fears and anxieties* (see McGlyn & Rose, 1998). The main cognitions that seem to characterize social phobias are interpersonal threat along with beliefs that positive interpersonal feedback is incorrect (Kendall & Treadwell, 2007; Sewitch & Kirsch, 1984). The importance of a cognitive assessment of social phobias is underscored by research suggesting that cognitive deficits and distortions are more important in causing and maintaining the difficulty than deficits in social skills (Heimberg, 1994). Those with social phobia are more likely to recall negative information, fear social embarrassment, interpret ambiguous feedback negatively, underestimate their own performance, expect more negative evaluations from others, and have more negative self-statements before interactions (Breitholtz et al., 1999; Cacioppo, Glass, & Merluzzi, 1979; Hope & Heimberg, 1993). Assessment of the relative rate of occurrence of each of these cognitive areas can inform specific treatment suggestions relating to which processes need to be modified. Some of the most frequently used instruments in the cognitive assessment of social phobias are the Social Avoidance and Distress Scale (Watson & Friend, 1969), the Fear of Negative Evaluation Scale (FNE; Watson & Friend, 1969), and the Social Interaction Self-Statement Test (Glass, Merluzzi, Biever, & Larsen, 1982). Many of the self-statements described by research on social phobias and measured by tests such as the Social Interaction Self-Statement Test are quite similar to the ones described by A. T. Beck (1967) as characteristic of depression. These similarities raise the still-unresolved issue of whether specific irrational beliefs are related to specific disorders or whether there is a nonspecific (yet generally negative) effect of irrational beliefs (see Heimberg, 1994). Although less work has been done on generalized anxiety, two frequently used tests are the Irrational Beliefs Test (R. Jones, 1969) and the Rational Behavior Inventory (Shorkey, Reyes, & Whiteman, 1977). The many versions of the Fear Survey Schedule (Wolpe & Lang, 1964, 1969, 1977) and the Fear Survey Schedule for Children (Ollendick, 1978, 1983) do not measure specific cognitions related to fear, but they are both frequently used and quite useful in detailing the various categories of fear a client experiences.

Several strategies have been used in the assessment of *eating disorders* based on the observations that this class of disorder involves considerable cognitive distortions (Mizes & Christiano, 1994). Some authors have taken a previously developed scale, such as the Cognitive Error Questionnaire (Lefebvre, 1981), and modified it to evaluate the cognitive distortions specific to eating disorders (Dritschel, Williams, & Cooper, 1991). The Eating Attitudes Test (Garner & Garfinkel, 1979) and the Bulimia Test—Revised (Thelen, Farmer, Wonderlich, & Smith, 1991) both have strong psychometric properties and focus primarily on cognitions related to eating and weight control. A further strategy is to have persons with eating disorders monitor their self-statements in their natural environments (Zotter & Crowther, 1991). The value of such strategies is the indication that cognitive behavioral instruments can be tailored

toward specific disorders and the information derived from these strategies has direct relevance for treatment, as it provides clinicians with specific cognitions to work with.

The area that has dominated the assessment of *social skills* has been assertiveness. Such assessment typically rates not only cognitions related to assertive behavior but also specific behaviors and skills. A wide variety of self-report inventories has been developed, including the Wolpe-Lazarus Assertion Inventory (Wolpe & Lazarus, 1966), the Conflict Resolution Inventory (McFall & Lillesand, 1971), the Gambrill Assertion Inventory (Gambrill & Richey, 1975), and the Bakker Assertiveness Inventory (Bakker, Bakker-Rab dau, & Breit, 1978). However, the RAS (Rathus, 1973) has been the most extensively used, and relevant normative data are available for normal college students (Quillan, Besing, & Dinning, 1977) as well as for psychiatric populations (Rathus & Nevid, 1977). Respondents are requested to rate, on a 6-point scale, how descriptive each statement is of them. A +3 indicates that the statement is “very uncharacteristic of me” and a –3 indicates that it is “very characteristic.” In addition to the original 30-item schedule, two other versions have been developed for special populations. The modified RAS (MRAS; Del Greco, Breitbach, & McCarthy, 1981) was developed for young adolescents. A simplified version is available that requires a minimum 6th-grade reading level, in contrast to the 10th-grade reading level required for the regular version (SRAS; McCormick, 1984). Additional, nonassertiveness social skills inventories include the Survey of Heterosexual Interactions (Twentyman & McFall, 1975) and the Stanford Shyness Survey (Zimbardo, 1977).

Assessing *marital relationships* involves gathering information about a wide range of behaviors with a particular focus on the strengths and weaknesses of the relationship, goals for change, and attempts the couple has made to change in the past (see Birchler & Fals-Stewart, 2006). Much of this information can and should be obtained through a careful interview. Areas related to cognitive assessment are the differing perceptions of each spouse, the perceived causes (attributions) for why the persons act in certain ways, expectations for future behavior, assumptions about relationships (roles, scripts), and standards by which the relationship is judged. Many of these areas can be evaluated through the use of cognitive self-report inventories. Some of the more frequent and well-researched instruments are the Relationships Beliefs Inventory (Eidelson & Epstein, 1982), the Dyadic Attribution Inventory (Baucom, Sayers, & Duhe, 1989), the Marital Attitude Survey (Pretzer, Epstein, & Fleming, 1992), the Relationship Attribution Measure (Fincham & Bradbury, 1992), and the Specific Relationship Standards (Baucom, Epstein, Rankin, & Burnett, 1996).

Self-efficacy has received considerable interest, particularly because it has been related to a variety of different predictions relevant to treatment (Bandura, 1986). Assessment is usually accomplished by simply having clients rate the degree to which they believe they are able to accomplish a certain skill or goal (e.g., stop smoking). Useful distinctions should be made between the level of strength of self-efficacy and its generalizability from one situation to another. Because some question exists regarding the degree to which self-efficacy can be related from one situation to the next, specific measurements are often used for different areas (e.g., depression, assertion, smoking, etc.). A person having a high level of self-efficacy is likely to have positive expectations about his or her effectiveness to judge and deal effectively with situations. Self-efficacy is developed as a result of the attainments someone has achieved in the past, vicarious

(observational) experiences, verbal persuasion, and physiological states. An assessment of self-efficacy is especially important in understanding the antecedent and retrospective accounts of the effect and quality of problem behavior. Relative levels of self-efficacy have been found to predict a wide number of variables, including general therapy outcome, treatment of smoking, and relapse rate from self-regulatory training (see Bandura, 1997).

Imagery has frequently been observed to relate to a person's presenting problem in the form of fantasies, daydreams, and different dreaming states. A depressed person may continually repeat images of being criticized, the anxious person might replay scenes of danger, and the paranoid person might frequently review images of persecution. Knowing a person's relative ability to produce and control images may be important in predicting response to treatment that requires the formation of images, such as systematic desensitization, covert desensitization, covert aversive conditioning, and certain types of relaxation procedures (Sheikh, 2003). Extensive experimental work has been conducted on imagery in areas like the different dimensions of imagery (Parks, 1982), differences between waking and nonwaking imagery (Cartwright, 1986), and the effects of conscious and unconscious images on behavior (Horowitz, 1985). Persons wishing to assess both clinical imagery and other aspects of cognitions might use one or several of the strategies described below that have been developed to assess cognitions.

Recording Cognitions

In addition to the many self-report inventories available, a number of strategies have been developed for recording cognitions in a less structured manner. Parks and Hollon (1988) listed and summarized the methods used by previous and subsequent researchers:

Thinking Aloud. Clients are requested to verbalize their ongoing thoughts, with these verbalizations usually extending for 5 to 10 minutes (Lodge et al., 2000). A similar technique is free association, in which the client is asked to simply say whatever comes to mind rather than report on his or her ongoing inner thoughts. A potential problem is that the procedure may feel unnatural and, therefore, provide a sample different from normally occurring internal thoughts. Also, the client may have no opportunity to verbalize competing thoughts with the result that the reported thoughts will most likely be a limited portion of total cognitions. In addition, clients may not report everything honestly. A factor that is likely to make the verbally reported thoughts different from actual ongoing processes is that, typically, people change the topic of ongoing internal dialogues every 5 to 6 seconds, whereas verbal reports of these dialogues may have topic changes only on the average of every 30 seconds.

Private Speech. Sometimes children's cognitions can be assessed by paying close attention to barely audible speech they make while engaged in various activities. It is believed that these private verbalizations are closely aligned to inner thoughts.

Articulated Thoughts. Clinicians may wish to create structured situations or simulations that parallel the problems the client reports. For example, an analog situation may be created that demands the client to be assertive or be exposed to criticism or phobic stimuli (Rosqvist et al., 2006). The person can then be asked to articulate the thoughts he or she is experiencing during these situations. Typical thoughts can be noted and inferences made regarding how they relate to the problem behaviors.

Production Methods. Instead of asking clients to articulate their thoughts during a simulation, an actual naturalistic situation (e.g., criticism, phobic stimuli, etc.) can occur, with clients then noting and recording the typical thoughts they have related to these situations. As such, these methods might also be referred to as in vivo self-reports.

Endorsement Method. The client might be presented with either a standardized (e.g., Irrational Beliefs Test, Cognitive Bias Questionnaire) or an informally developed list of items and then be requested to rate frequency of occurrence, strength of belief, and how the item might be uniquely represented in the person's cognitions. These items might include ratings of the frequency of such thoughts as "What's the use?" or "I can't do anything right." Potential difficulties with this technique are the effects of the demand characteristics of the situation and social desirability. An underlying and questionable assumption behind the technique is that the relevant cognitions are in the client's conscious awareness.

Thought Listing. Instead of developing a continuous description of ongoing thoughts, clients might be asked simply to summarize their relevant thoughts. These thoughts might be elicited by a specific stimulus, problem area, or merely attending to or anticipating a stimulus.

Thought Sampling. A sample of a person's thoughts might be obtained by setting a prompt (e.g., a beep on a smartphone), then having the client describe the thoughts he or she was having just before the interruption by the prompt.

Event Recording. The client might be asked to wait until a relevant event occurs (e.g., hand washing for persons with obsessive-compulsive disorder), at which point the thoughts related to these events are written down. Instead of merely waiting for a problem or spontaneously occurring behavior, a client might be asked to describe the thoughts related to the expression of new and desired behaviors, such as assertion. The relevant thoughts about these behaviors might then be used to increase the likelihood of their continued occurrence.

Psychophysiological Assessment

A complete understanding of the person involves an assessment of not only behavioral, affective, and cognitive modes, but also of the ways these interact with and are dependent on physiological functioning. Such psychophysiological assessments have recently become easier to make because of increased interest and knowledge regarding instrumentation (magnetic resonance imaging [MRI], electronics, computers), operant conditioning of behaviors that at one time were considered involuntary, physiological and neurochemical aspects of behavior, and behavioral medicine (Larkin, 2006). The most

frequently assessed physiological responses are heart rate, blood pressure, skin temperature, muscle tension, vasodilation, galvanic skin response (GSR), and brain activity as measured by electroencephalograms (EEGs). Functional magnetic resonance imaging (fMRI) has become popular as well, at least in the research realm (see Aue, Lavelle, & Cacioppo, 2009). By quantifying data gathered through these areas, psychological problems can be translated into more precise physiological indices.

One of the first relevant studies to relate psychological and physiological modes indicated that fear and anger had different physiological responses in blood pressure and skin conductance (Ax, 1953). This result suggested that these and other psychological variables might be measured in ways other than through self-report inventories. More recently, it has been found that patients with obsessive-compulsive disorder had greater activation of the orbitofrontal region of the cortex (K. E. Anderson & Savage, 2004). A further representative area of research has involved the relationship between different personality variables and psychophysiological measurement. Clients with antisocial personalities have been found to have lower skin conductance than those with anxiety disorders (Lorber, 2004). Polygraph testing to detect lying, while still in extensive use, has not been found to be particularly effective at detecting those who lie from those who are telling the truth (Iacano & Patrick, 2006). In contrast, greater promise has been demonstrated differentiating true from faked memory loss using event-related potentials (ERPs). Physiological baseline measures for an area such as anxiety can and have been used to monitor the effectiveness of treatment for social phobias, generalized anxiety disorders, and obsessive-compulsive disorders (Larkin, 2006). Although most of the previously mentioned studies represent very general correlations among such variables as emotions, personality, behavioral disorders, and outcome assessment, they show considerable potential for future assessment, should these measures become more refined.

In addition to the usual knowledge relating to psychological assessment, clinicians who obtain and interpret psychophysiological data must have knowledge in anatomy; electronics; and the physiology of cardiovascular, musculoskeletal, neurological, respiratory, electrodermal, ocular, and gastrointestinal response systems. This extensive background is particularly important because instrumentation presents a number of special problems. A variety of confounding factors may be present, such as the effect of slowing respiratory rate to alter cardiac output or the effect of eye roll on measured brain activity. Filters might be necessary to exclude noise in the system. The techniques are also intrusive, thereby making the situation artificial. As a result, it may not be correct to generalize to outside aspects of the client's life or between different response modes. A wide variety of difficulties may arise regarding meaningful psychological interpretations based on physiological data. In the future, the development of better instruments and improved methods of computer analysis is likely to increase the utility of psychophysiological assessment and overcome many of these difficulties.

RECOMMENDED READING

- Haynes, S. N., & Heiby, E. M. (Eds.). (2004). *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3). Hoboken, NJ: Wiley.
- Haynes, S. N., O'Brien, W., & Kaholokula, J. (2011). *Behavioral assessment and case formulation*. Hoboken, NJ: Wiley.

- Hersen, M. (Ed.). (2005a). *Clinician's handbook of adult behavioral assessment*. New York, NY: Elsevier.
- Hersen, M. (Ed.). (2005b). *Clinician's handbook of child behavioral assessment*. New York, NY: Elsevier.
- Hersen, M., & Bellack, A. S. (Eds.). (2002). *Dictionary of behavioral assessment techniques* (2nd ed). New York, NY: Pergamon Press.
- McLeod, B. D., Jensen-Doss, A., & Ollendick, T. H. (Eds.). (2013). *Diagnostic and behavioral assessment in children and adolescents: A clinical guide*. New York, NY: Guilford Press.
- Whitcomb, S., & Merrell, K. W. (2012). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.

WECHSLER INTELLIGENCE SCALES

The Wechsler intelligence scales are individually administered, composite intelligence tests in a battery format. They assess different areas of intellectual abilities and create a situation in which aspects of personality can be observed. The most recent versions (WAIS-IV and WISC-V) provide an overall, or “Full Scale” IQ (FSIQ) as well as specific index scores that can be calculated using various combinations of subtests. The Wechsler intelligence scales are considered to be among the best of all psychological tests because they have sound psychometric properties and produce information relevant to practitioners. As a result, they have become the most frequently used tests in clinical practice (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Camara, Nathan, & Puente, 2000; C. E. Watkins, Campbell, Nieberding, & Hallmark, 1995).

TESTING OF INTELLIGENCE: PROS AND CONS

The testing of intelligence has had a consistent history of misunderstanding, controversy, and occasional misuse (Bartholomew, 2006; Flanagan & Harrison, 2005; Weinberg, 1989). Criticisms have ranged from moral indictments against labeling individuals, to cultural bias, and even to accusations of flagrant abuse of test scores. Although valid criticisms can be made against testing intelligence, such procedures also have a number of advantages.

One of the main assets of intelligence tests is their accuracy in predicting future behavior. Initially, Alfred Binet was able to achieve a certain degree of predictive success with his scales, and, since that time, test procedures have become progressively more refined and accurate. More recent studies provide ample support that intelligence tests can predict an extremely wide number of variables. In particular, IQ tests are excellent predictors of academic achievement (Kaufman & Lichtenberger, 2006; Neisser et al., 1996) and occupational performance (J. Hunter & Schmidt, 1996; F. L. Schmidt & Hunter, 1998, 2004; Wagner, 1997) and are sensitive to the presence of neuropsychological deficit (Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Lezak, Howieson, Bigler, & Tranel, 2012). However, certain liabilities are also associated with these successes. First, intelligence tests can be used to classify children into stereotyped categories, which may limit their freedom to choose fields of study. Furthermore, IQ tests are quite limited in predicting nontest or nonacademic activity, yet sometimes they are incorrectly used to make these inferences (Snyderman & Rothman, 1987; Sternberg, 2003). It should also be stressed that intelligence tests are measures of a person’s current level of functioning and, as such, are best used for making short-term predictions. Long-term predictions, although attempted frequently, are less accurate because there

can be many uncontrolled, influencing variables. Similarly, even short-term academic placements made solely on the basis of an IQ score have a high chance of failure because all the variables that may be crucial for success are not and cannot be measured by a test. It can sometimes be tempting for test users to extend the meaning of test scores beyond their intended scope, especially in relation to the predictions they can realistically be expected to make.

In addition to predicting academic achievement, IQ scores have been correlated with occupation, ranging from highly trained professionals with mean IQs of 125, to unskilled workers with mean IQs of 87 (Reynolds, Chastain, Kaufman, & McLean, 1987). Correlations between job proficiency and general intelligence have been higher in predicting relatively more complex jobs (.58) than less demanding occupations (.23; F. L. Schmidt & Hunter, 2004). Schmidt and Hunter (2004) also reported moderately high correlations between general intelligence and success for managers (.53), salespersons (.61), and clerks (.54). For intellectually demanding tasks, nearly half the variance related to performance criteria can be accounted for by general intelligence (F. L. Schmidt & Hunter, 2004; F. L. Schmidt, Ones, & Hunter, 1992). The use of intelligence tests for personnel selection has demonstrated financial efficacy for organizations (F. L. Schmidt & Hunter, 1998). In addition, the accuracy of using IQ tests can be incrementally increased by combining the results with integrity tests, work samples, and structured interviews (F. L. Schmidt & Hunter, 1998, 2004).

Another important asset of intelligence tests, particularly the WAIS-IV and WISC-V, is that they provide valuable information about a person's cognitive strengths and weaknesses. They are standardized procedures whereby a person's performance in various areas can be compared with that of age-related peers. In addition, useful comparisons can be made regarding a person's own pattern of strengths and weaknesses. The WAIS-IV, WISC-V, and other individually administered tests provide the examiner with a structured context in which a variety of tasks can be used to observe the unique and personal ways the examinee approaches cognitive tasks. Through a client's interactions with both the examiner and the test materials, the examiner can gain an initial impression of the individual's self-esteem, behavioral idiosyncrasies, anxiety, social skills, and motivation while also obtaining a specific picture of intellectual functioning.

Intelligence tests often provide clinicians, educators, and researchers with baseline measures for use in determining either the degree of change that has occurred in an individual over time or how an individual compares with other persons in a particular area or ability. These distinctions may have important implications for evaluating the effectiveness of an educational program or for assessing the changing abilities of a specific student. In cases involving recovery from a head injury or readjustment following neurosurgery, it may be extremely helpful for clinicians to measure and follow the cognitive changes that occur in a patient. Furthermore, IQ assessments may be important in researching and understanding more adequately the effect on cognitive functioning of environmental variables, such as educational programs, family background, and nutrition. Thus, these assessments can provide useful information about cultural, biological, maturational, and treatment-related differences among individuals.

A criticism leveled at intelligence tests is that almost all have an inherent bias toward emphasizing convergent, analytical, and scientific modes of thought. Thus, a person who emphasizes divergent, artistic, and imaginative modes of thought may be at a

distinct disadvantage. Some critics have even stressed that the current approach to intelligence testing has become a social mechanism used by people with similar values to pass on educational advantages to children who resemble themselves. Not only might IQ tests tend to place creative individuals at a disadvantage, but also they are limited in assessing nonacademically oriented intellectual abilities (Gardner, 2006; Snyderman & Rothman, 1987; Sternberg, 2003). Thus, social acumen, success in dealing with people, the ability to handle the concrete realities of the individual's daily world, social fluency, and specific tasks, such as purchasing merchandise, are not measured by standard intelligence tests (Sternberg, 2003). More succinctly, people are capable of many more cognitive abilities than can possibly be measured on an intelligence test.

Misunderstanding and potential misuse of intelligence tests frequently occur when scores are treated as measures of innate capacity. IQ is not a measure of an innate fixed ability, nor is it representative of all problem-solving situations. It is a specific and limited sample, made at a certain point in time, of abilities that are susceptible to change because of a variety of circumstances. It reflects, to a large extent, the richness of an individual's past experiences. Although interpretation guidelines are quite clear in pointing out the limited nature of a test score, there is a tendency to look at test results as absolute facts reflecting permanent characteristics in an individual. People often want a quick, easy, and reductionist method to quantify, understand, and assess innate cognitive abilities, and the IQ score has become the most widely misused test score to fill this need.

An important limitation of intelligence tests is that, for the most part, they are not concerned with the underlying processes involved in problem solving. They focus on the final product or outcome rather than on the steps involved in reaching the outcome. They look at the "what" rather than the "how" (Embretson, Schneider, & Roth, 1986; E. Kaplan et al., 1999; Milberg, Hebben, & Kaplan, 1996). Thus, a low score on Arithmetic might result from poor attention, difficulty understanding the examiner because of disturbances in comprehension, dyscalculia, or low educational attainment. The extreme example of this "end product" emphasis is the global IQ score. When the examiner looks at the myriad assortment of intellectual abilities as a global ability, the complexity of cognitive functioning may be oversimplified to the point of being almost useless. The practitioner can apply labels quickly and easily, without attempting to examine the specific strengths and weaknesses that might make precise therapeutic interventions or knowledgeable recommendations possible. Such thinking detracts significantly from the search for a wider, more precise, and more process-oriented understanding of mental abilities.

A further concern about intelligence tests involves their limited ability to accurately assess minority groups with divergent cultural backgrounds. It has been stated that intelligence test content is biased in favor of European American, middle-class values. Critics stress that minorities tend to be at a disadvantage when taking the tests because of deficiencies in motivation, lack of practice, lack of familiarity with culturally loaded items, and difficulties in establishing rapport. Numerous arguments against using intelligence tests for the assessment and placement of minorities have culminated in legal restrictions on the use of IQ scores. However, traditional defenses of IQ scores suggest that they are less biased than has been accused (see the "Use with Diverse Groups" section later in this chapter). The issue certainly has not been resolved, but clinicians

should continue to be aware of this dilemma, pay attention to subgroup norms, and interpret minority group IQ scores cautiously.

Finally, many people feel that their IQs are deeply personal pieces of information. They would prefer that others, even a psychologist who is expected to observe confidentiality, not be allowed access to this information. This problem is further compounded when IQ scores might be given to several different persons, such as during legal proceedings or personnel selection.

Intelligence tests provide a number of useful and well-respected functions. They can: adequately predict short-term scholastic performance; assess an individual's relative strengths and weaknesses; predict occupational achievement; reveal important personality variables; and permit the researcher, educator, or clinician to trace possible changes in an individual or population. However, these assets are helpful only if the limitations of intelligence tests are adequately understood and appropriately taken into consideration. The tests are limited in predicting certain aspects of occupational success and nonacademic skills, such as creativity, motivational level, social acumen, and success in dealing with people. Furthermore, IQ scores are not measures of an innate, fixed ability, and their use in classifying minority groups has been questioned. Finally, there has been an overemphasis on understanding the end product of cognitive functioning and a relative neglect in appreciating underlying cognitive processes.

HISTORY AND DEVELOPMENT

During the 1930s, Wechsler began studying a number of standardized tests and selected 11 different subtests to form his initial battery. His search for subtests was in part guided by his conception that intelligence is global in nature and represents a part of the greater whole of personality. Several of his subtests were derived from portions of the 1937 revision of the Stanford-Binet (Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary). The remaining subtests came from the Army Group Examinations (Picture Arrangement), Koh's Block Design (Block Design), Army Alpha (Information, Comprehension), Army Beta (Digit Symbol-Coding), Healy Picture Completion (Picture Completion), and Pinther-Paterson Test (Object Assembly). These subtests were combined and published in 1939 as the Wechsler-Bellevue Intelligence Scale. The Wechsler-Bellevue had a number of technical deficiencies primarily related to both the reliability of the subtests and the size and representativeness of the normative sample. Thus, it was revised to form the Wechsler Adult Intelligence Scale (WAIS) in 1955; another revised edition (WAIS-R) was published in 1981. The 1981 revision was based on 1,880 individuals who were generally representative of the 1970 census and categorized into nine different age groups.

In 1997, the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III) replaced the earlier WAIS-R. The primary reason for the revision was to update the norms. Additional reasons included extending the age range, modifying items, developing a higher IQ “ceiling” and “floor,” decreased reliance on timed performance, developing index/factor scores, creating linkages to other measures of cognitive functioning/achievement, and extensive testing of reliability and validity. Despite these changes, many of the traditional features of the WAIS-R were maintained, including the six

verbal subtests and the five performance subtests. Maintaining these clusters of subtests still enabled practitioners to calculate the Full Scale, Verbal, and Performance IQs. An added feature of the WAIS-III was the inclusion of three new subtests, which enabled the calculation of four index scores. Thus, the WAIS-III was not merely a renormed face-lift; it also enabled the clinician to do more with the different test scores, such as being able to assess persons with either greater age or IQ ranges, linking scores with the Wechsler Memory Scales, and calculating both IQ and index/factor scores.

The Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) is the most recent revision of the evolving Wechsler intelligence scales for adults (Wechsler, 2008a). The general purpose of the revision was to update norms, improve floors and ceilings, improve psychometric properties, reduce testing time, and conorm it with the Wechsler Memory Scale—Fourth Edition (WMS-IV; see Table 5.1) and the Wechsler Individual Achievement Test—Second Edition (WIAT-II, although now that test is on its third edition, the WIAT-III). One of the most obvious changes has been the elimination of the time-honored verbal versus performance IQ. Instead, the WAIS-IV uses the traditional Full Scale IQ along with four index scores (Verbal Comprehension, Working Memory, Perceptual Reasoning, and Processing Speed). The major rationale for the elimination of the Verbal-Performance IQs is that they are not pure measures but typically combine a number of different abilities. For example, the Verbal IQ included measures of verbal abilities as well as working memory. Thus, it was not a unitary measure of an ability. In contrast, relying on the four index scores ensures that relatively pure, theoretically sound measures of abilities have been made. This reliance on a Full Scale IQ plus the four indexes paralleled similar development for the WISC-IV (Wechsler, 2003a, 2003b), however the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014a) broadened the structure out to include both the Full Scale IQ and *five* index scores. In addition, the WAIS-IV

Table 5.1 Major Changes on the WAIS-IV

Elimination of Verbal and Performance IQs
Updated norms
15 subtests (versus 14 on the WAIS-III)
Computation of FSIQ and indexes based on 10 core subtests
General Ability Index (optional index that combines Verbal Comprehension and Perceptual Reasoning)
3 newly developed subtests (Visual Puzzles, Figure Weights, Cancellation)
Deletion of 2 subtests (Object Assembly, Picture Arrangement)
Organization of subtests into core and supplemental
Renaming of the Perceptual Organization Index to the Perceptual Reasoning Index
Inclusion of process scoring options for Block Design, Digit Span, Letter-Number Sequencing
Potential for shortened administration using only the 10 core subtests (for FSIQ + indexes)
Greater attention to floor and ceilings
Normed linkages with the Wechsler Memory Scale–IV
Upgrade kit for specialist neuropsychologists and geropsychologists

includes an optional General Ability Index that combines the Verbal Comprehension and Perceptual Reasoning index scores. An upgrade of the WAIS-IV for neuropsychologists and geropsychologists became available in 2009 (*WAIS-IV/WMS-IV Advanced Clinical Solutions*, Pearson, 2009a).

A further feature of the WAIS-IV has been the deletion, addition, and revision of subtests. In addition, subtests have been organized according to core and supplemental subtests. The core subtests are used to develop the index scores (see Table 5.2). However, if a core subtest is “spoiled” (i.e., made invalid), if practitioners are unable to administer a core subtest, or if performance on a core subtest is so divergent from performance on the other subtests that the practitioner suspects something went wrong, it can be replaced with one of the supplemental subtests. Supplemental subtests can also be administered to find additional information regarding a client’s level of functioning. For example, the new Cancellation subtest might be added to Symbol Search and Coding to add further information related to a client’s ability to process information rapidly; or it may be used on its own to evaluate the client’s ability to quickly attend to information in the face of multiple distractors. Many of the subtests have undergone revisions to enhance clarity of instructions, refine scoring rules, change stimuli, and include different items.

New norms have been developed for the WAIS-IV derived from 2,200 persons between the ages of 16 and 90 stratified according to sex, education, ethnicity, and geographical region. These norms closely correspond to the 2005 U.S. census data. Whereas 200 examinees were included for the age bands between 16 and 60, only 100 examinees were included for the age bands between 70 and 90. The WAIS-IV was conormed with the WMS-IV and WIAT-II. Norms and patterns of responses have been developed for special groups, including mild cognitive impairment, borderline intellectual functioning, traumatic brain injury, Alzheimer’s disease, attention-deficit/hyperactivity disorder (ADHD), reading disorder, math disorder, autism, asperger’s syndrome, and depression.

The original Wechsler-Bellevue Scale was developed for adults, but in 1949, Wechsler developed the Wechsler Intelligence Scale for Children (WISC) so that children from the age of 5 years 0 months could be assessed in a similar manner. Easier items,

Table 5.2 Organization of WAIS-IV Subtests

Index	Core subtests	Supplemental subtests
Verbal Comprehension	Similarities Vocabulary Information	Comprehension
Perceptual Reasoning	Block Design Matrix Reasoning Visual Puzzles	Figure Weights Picture Completion
Working Memory	Digit Span Arithmetic	Letter-Number Sequencing
Processing Speed	Symbol Search Coding	Cancellation

designed for children, were added to the original scales and standardized on 2,200 European American boys and girls selected to be representative of the 1940 census. However, some evidence shows that Wechsler's sample may have been overrepresentative of children in the middle and upper socioeconomic levels. Thus, ethnic minorities and children from lower socioeconomic levels may have been necessarily penalized when compared with the normative group. The WISC was revised in 1974 and standardized on a new sample that was more accurately representative of children in the United States. The WISC-III (Wechsler, 1991) was released in 1991; its major change was the inclusion of four factor/index scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed). The new Processing Speed factor involved the inclusion of a new Symbol Search subtest along with the older Coding subtest. As with the earlier WISC-R, the standardization and reliability were excellent. The scales were standardized on 2,200 children between the ages of 6 and 16 who closely matched the 1988 census. The sample consisted of 100 boys and 100 girls for each of the different age groups.

The WISC-IV (Wechsler, 2003a) was noteworthy in that it contained more changes than any other previous edition (see Table 5.3). The most obvious change was the elimination of the time-honored Verbal and Performance IQ. There was instead a greater reliance on interpretation using a combination of the four index scores along with the global Full Scale IQ (see Table 5.4). The indexes were also refined by the inclusion of five new subtests (and the deletion of Picture Arrangement, Object Assembly, and Mazes). Completely new norms for the WISC-IV were developed

Table 5.3 Major Changes on the WISC-V

Perceptual Reasoning Index split into Visual Spatial Index and Fluid Reasoning Index
Updated norms
18 subtests (versus 15 on the WISC-IV)
Computation of IQ and Indexes based on 7 core subtests
Introduction of 5 new WISC-V subtests:
3 newly developed subtests (Picture Span, Naming Speed, Symbol Translation) plus downward extensions of 2 WAIS-IV subtests (Visual Puzzles, Figure Weights)
2 deleted subtests (Word Reasoning, Picture Completion)
New organization of subtests into core and supplemental
Introduction of 3 new Ancillary Index Scores (Quantitative Reasoning, Auditory Working Memory, Nonverbal) and 3 new Complementary Indexes (Naming Speed, Symbol Translation, Storage and Retrieval)
Reduced administration time for 10 subtests needed for the FSIQ (48 minutes) or for the 5 index scores (65 minutes)
Potential for shortened administration using only the 7 core subtests (for FSIQ + indexes)
Revised items with attention to cultural portability, increased floor–ceiling range, reduction of confusion as basis for incorrect responses (e.g., eliminated rhyming letters and numbers on Letter-Number Sequencing)
Normed linkages with the Wechsler Individual Intelligence Test–III

Table 5.4 Organization of WISC-IV Indexes and Subtests

Index	Core subtests	Supplemental subtests
Verbal Comprehension	Similarities Vocabulary	Information Comprehension
Visual Spatial	Block Design Visual Puzzles	
Fluid Reasoning	Matrix Reasoning Figure Weights	Picture Concepts Arithmetic
Working Memory	Digit Span Picture Span	Letter-Number Sequencing
Processing Speed	Coding Symbol Search	Cancellation

that closely represented the U.S. census. A further potentially useful feature was the publication of the WISC-IV Integrated (Wechsler et al., 2004), which allowed for 12 additional procedures that enabled specialty practitioners the option of analyzing the underlying processes clients go through when making their responses (see McCloskey & Maerlander, 2005).

In 2014, the WISC-V (Wechsler, 2014a) came out with further revisions to the test. The most notable difference is the separation of the previous Perceptual Reasoning Index (PRI) into two indexes, the Visual Spatial Index (VSI) and the Fluid Reasoning Index (FRI). This change adds a fifth index to the structure of what underlies the Full Scale IQ and was based on a better understanding that the PRI was tapping two different skill constructs. Additional changes include the deletion of several subtests (Word Reasoning, which had an extremely high correlation with the information subtest, and Picture Completion, which did not seem to assess visual ability well enough), the addition of several subtests (Visual Puzzles, Figure Weights, Picture Span, and the supplemental subtests Naming Speed and Symbol Translation), and the revision of content on most of the remaining subtests (including changes to make the subtests clearer, less susceptible to errors that are not due to the skill being tapped, and more culturally portable). The WISC-V maintained the ability to interpret the Full Scale IQ, the five individual indexes, and to calculate additional clustered scales, such as the General Ability Index (GAI), the Cognitive Proficiency Index (CPI), and the Quantitative Reasoning.

One of the motivations for the WAIS-IV and WISC-V revisions was to update the instruments' theoretical foundations. To a certain extent, this has been done. The importance of fluid intelligence has been reflected in the introduction of subtests that assesses this area of intellectual functioning (Matrix Reasoning, Picture Concepts, and Figure Weights). In addition, the concepts of working memory and processing speed have been incorporated and refined. The result has been changes and refinements in the subtests and psychometric properties included in the Working Memory and Processing Speed indexes. It should also be noted that the factor structure of the more recent revisions of the Wechsler intelligence scales has resulted in a *de facto* theory of

intelligence defined by the Full Scale IQ in combination with the functions measured by the four/five indexes. However, the WAIS-IV was not organized around a single specific theory of intelligence. In contrast, revisions of the K-ABC, Stanford Binet—5, and the Woodcock Johnson—III, each of which occurred in the first few years of the 21st century, were closely aligned to the Cattell-Horn-Carroll (CHC) theory of intelligence, and the WISC-V aligns more closely with it as well, with the introduction of the Fluid Reasoning Index. Some have stated that the Wechsler intelligence scales have been overburdened by their traditions, resulting in a failure to make major adaptations to evolving knowledge related to intelligence (Flanagan & Kaufman, 2009).

In 1967, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was first published for the assessment of children between the ages of 4 and 6 years 6 months. Just as the WISC is a downward extension of the WAIS, so the WPPSI is generally a downward extension of the WISC in which easier but similar items are used. Although most of the scales are similar in form and content to the WISC, a number are unique to the WPPSI. The WPPSI was revised in 1989 (WPPSI-R; Wechsler, 1989), again in 2002 (WPPSI-III; Wechsler, 2002c), and again in 2012 (WPPSI-IV; Wechsler, 2012).

RELIABILITY AND VALIDITY

WAIS-IV Reliability and Validity

The reliabilities for the WAIS-IV are generally quite high (Wechsler, 2008a). An area of note is that average split-half reliability for the Full Scale IQ is extremely high at .98 (Wechsler, 2008a). Average split-half reliabilities for the other combined or “composite” scores are only slightly lower and range from a high of .96 for the Verbal Comprehension index to a low of .90 for Processing Speed. Average split-half reliability for the subtests ranged from excellent (i.e., Vocabulary $r = .94$, Digit Span $r = .93$) to an acceptable .78 for Cancellation. All subtests except Cancellation were above .81. These good to excellent reliabilities were found not only for the standardization sample but also among various clinical populations (e.g., those with brain injury, ADHD, Alzheimer’s disease). The average standard error of measurement (SEM) for various WAIS-IV scores indicates a small band of error (i.e., Full Scale IQ = 2.16, Verbal Comprehension Index = 2.12, Processing Speed Index = 4.24). Calculating the SEM for each of the composite scores is a standard procedure on the WAIS-IV Record Form so that it is reported for all examinees. The average test-retest reliabilities (8–82 days, $M = 22$ days) for the Full Scale IQ was quite high ($r = .96$), and the composite scores were similarly high, ranging from .96 for Verbal Comprehension to a low of .87 for Processing Speed. These reliabilities are among the best for any test available and, in almost all cases, represent a slight improvement over the WAIS-III.

While these test-retest reliabilities indicate a high degree of temporal stability, there is still some degree of improvement on retesting because of practice effects. Improved performance due to retesting is important to understand; clinicians need to know when to attribute an increase in scores to practice effects and when this might indicate actual clinical improvement. Over the retesting interval (8–82 days, $M = 22$ days), the Full Scale IQ was found to increase by 4.7 points. The lowest increase was for Verbal Comprehension (2.5 points), followed by Working Memory (3.1), Perceptual

Reasoning (3.9), and Processing Speed (4.4). These increases are not only statistically significant but may have clinical significance when making inferences about the extent to which real improvement/deterioration has occurred for a particular client. Thus, a client who has a Perceptual Reasoning increase of 4 points on retesting may not really be improving in everyday functions but merely demonstrating practice effects. A difference of 15 points on the earlier WAIS-III Full Scale IQ (for ages 16 to 54) was found to be necessary to infer that there had been an actual improvement in abilities (Kaufman & Lichtenberger, 2006). Research with the WAIS-R indicated that these practice effects can occur up to 9 months later, even among patients with head injuries. However, retest gains have also been found to diminish with advancing age (J. Ryan, Paolo, & Brungardt, 1990; Wechsler, 2008a, 2008b).

Because extensive validity studies exist for the WAIS-III, one of the most important steps in WAIS-IV validation was to determine the comparability between the two tests. As expected, correlations were found to be quite high. The WAIS-IV and WAIS-III Full Scale IQ correlation was .94 (Wechsler, 2008b). The four indexes were similarly high, ranging from .91 for Verbal Comprehension to .84 for Perceptual Reasoning/Perceptual Organization. This finding suggests that the WAIS-IV measures essentially the same constructs as the WAIS-III. Noteworthy high correlations between WAIS-IV and WAIS-III subtests were .90 for Information, .87 for Vocabulary, and .85 for Coding. In contrast, a relatively low correlation was found for Picture Completion (.65). Correlations between the WAIS-IV and WISC-V for a group of 16-year-olds were quite high (Full Scale IQ = .89, Verbal Comprehension = .83, Perceptual Reasoning/Visual Spatial = .83, Perceptual Reasoning/Fluid Reasoning = .62, Working Memory = .76, Processing Speed = .83; Wechsler, 2014b). As Fluid Reasoning includes only one subtest overlap with Perceptual Reasoning and includes different skills, the slightly lower correlation was expected. The correlations between the WAIS-IV and WMS-III were .61 for the Working Memory Index. A similar .59 correlation was found between the Full Scale IQ and the WMS-IV General Memory Index. These moderate correlations are expected, given that the WAIS-IV and WMS-IV measure somewhat different but still overlapping constructs. Correlations between achievement were, as expected, in the moderate to high range (WAIS-IV Full Scale IQ and Wechsler Individual Achievement Test—II Total Achievement = .88).

The WAIS-IV has also been found to produce expected patterns of correlation with a number of additional standard ability measures (Wechsler, 2008b). The Delis-Kaplan Executive Functioning System (Delis, Kaplan, & Kramer, 2001) is a series of subtests that measure various aspects of a client's ability to initiate, plan, and flexibly monitor their behavior. Representative correlations were a .22 between Perceptual Reasoning and the ability to fluidly produce the names of classes of objects (Category Fluency) and a correlation of .77 with Full Scale IQ and the ability to flexibly and rapidly connect combinations of letters and numbers (Trail Making). The California Verbal Learning Test—II (Delis, Kramer, Kaplan, & Ober, 2000) measures how well a person can recall lists of words that are read to them. Correlations between the WAIS-IV Full Scale IQ and a series of trials on learning word lists ranged from .48 to .32 (Wechsler, 2008b). A final representative test is the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 2012), which measures various domains

of cognitive functioning (see Chapter 12). Correlations between the WAIS-IV Full Scale IQ and the RBANS Total Score was .75. The WAIS-IV Index scores and RBANS Total Score correlations were slightly lower, ranging from a high of .72 for Perceptual Reasoning to a low of .54 for Processing Speed. This overview of WAIS-IV correlations with various standardized measures provides strong support for WAIS-IV validity.

Factor analysis of the WAIS-IV has supported the presence of *g* in that most subtests correlate with each other as well as with the Full Scale IQ at least to a moderate extent (Wechsler, 2008b). Dividing subtests into four indexes is supported by current theories of intelligence as well as factor-analytic procedures (Wechsler, 2008b). However, Arithmetic was found to load on both the Verbal Comprehension as well as the Working Memory factors. This is consistent with conceptualizations of Arithmetic in that it involves both verbal abilities and working memory. In addition, the Figure Weights subtest was found to load highly on both Perceptual Reasoning and Working Memory. Again, these high loadings are expected, given that the Figure Weights subtest involves visual reasoning, but this reasoning is related to quantitative manipulations. The relation of these subtests to multiple factors is reflected in the WISC-V reorganization of them onto a separate skill domain, Fluid Reasoning.

Various clinical populations have patterns of deficits in learning, cognition, and memory (see Wechsler, 2008b). It is thus to be expected that the WAIS-IV would be sensitive to these patterns. This finding was somewhat supported in that the mean WAIS-IV Full Scale IQ ($M = 81.2$) and index scores for patients with Alzheimer's disease were low compared with their age-related peers. Comparisons among the index scores indicated differential cognitive abilities in that the mean Verbal Comprehension Index was relatively higher (86.2) than Processing Speed (76.6). However, it would have been expected that the Working Memory Index would have been somewhat lower than the mean of 84.3, given the considerable memory complaints among this population. Patients with traumatic brain injuries had a somewhat similar pattern in that verbal abilities were relatively spared (relatively higher Verbal Comprehension mean of 92.1) when compared with a relatively lower Processing Speed (80.5). This finding indicates that the WAIS-IV is sensitive to the difficulties these patient populations have with rapidly processing and consolidating information.

The mean WAIS-IV Full Scale IQ (96.9) for clients diagnosed with ADHD did not differ substantially from the standardization sample. In addition, their mean Working Memory Index score (94.7) was only slightly lower than their Verbal Comprehension mean scores (100.9). This finding suggests that the WAIS-IV is not particularly sensitive to the attentional problems of this group, perhaps because the WAIS-IV is administered in a structured testing situation with few distractions. In contrast, real-world environments are likely to have multiple concurrent attentional demands that would be much more difficult for these clients to ignore. Subjects diagnosed with reading-related learning disabilities were found to have mean working memory scores (88.9) that were significantly below the WAIS-IV standardization sample (101.1). Individuals with learning disabilities characterized by mathematical difficulties similarly had the greatest difficulty with working memory (84.1) when compared with matched controls derived from the standardization sample (98.7). This finding reflects the common difficulties related to tasks requiring short-term memory

and attention. These examples of research described in the *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008b) suggest that the WAIS-IV is sensitive to the types of cognitive difficulties found among various patient groups.

WISC-V Reliability and Validity

Reliability on the WISC-IV is generally excellent. Internal consistency reported in the *Technical and Interpretive Manual* (Wechsler, 2014b) for the Full Scale IQ ranges from .96 to .97 ($M = .96$). The mean internal consistencies for the individual index scores range from .88 (for Processing Speed) to .93 (for Fluid Reasoning). The mean (across ages) internal consistencies for the 16 subtests range from a low of .81 for Symbol Search to a high of .94 for Figure Weights. Test-retest reliability (average 26-day interval) for the Full Scale IQ is .92. The five index test-retest reliabilities range from a low of .75 for Fluid Reasoning to a high of .94 for Verbal Comprehension. Average test-retest stability for the subtests range from a low of .71 for Picture Concepts to a high of .90 for Vocabulary. All subtest stability except Picture Concepts and Matrix Reasoning (.78) are .80 or higher.

The WISC-IV had well-substantiated validity, and it can be cautiously assumed that much of this research can be transferred to the WISC-V. This idea is partially supported in that more than half of the subtests on the WISC-V were retained (with similar content) from the WISC-IV. In addition, there are moderate to high correlations between the WISC-IV and WISC-V Full Scale IQs, index scores, and subtests (i.e., Full Scale IQ = .86, Verbal Comprehension = .85, Perceptual Reasoning/Visual Spatial = .66, Perceptual Reasoning/Fluid Reasoning = .63; Working Memory = .65, Processing Speed = .71). Criterion validity has been performed on several Pearson Assessment tests with generally favorable results. For example, the WISC-V indexes and appropriate index scores on the Kaufman Assessment Battery for Children–II (KABC-II) showed good convergence: Verbal Comprehension and Knowledge/Gc (crystallized ability) correlated .74; Visual Spatial and Nonverbal Index correlated .60; Working Memory and Mental Processing Index correlated .65; and Fluid Reasoning and Fluid-Crystallized Index correlated .63. Correlations with the Behavior Assessment System for Children–2 (BASC-2) Parent Rating Scale were low to nonsignificant, which is what would be expected, given that they are theoretically assessing different variables. The *Technical and Interpretive Manual* provides relationships with other measures as well as studies of nine different special groups. For example, a group with traumatic brain injury had a mean Processing Speed score of 84 and a mean Full Scale IQ of 83. In contrast, children who were considered to be intellectually gifted had mean Full Scale IQ scores of 127.5, with mean Verbal Comprehension scores of 127.7.

Confirmatory factor analyses generally supported the five-factor model reported in the *Technical and Interpretive Manual* and reflected in the five index scores; however, the developers utilized higher-order models instead of bifactor models, which some researchers have argued is not appropriate (Canivez, 2014; Canivez & Watkins, in press). The theoretical development of the WISC-V was predicated on coming closer to the CHC theory of intelligence (see Flanagan & Kaufman, 2009; Keith, Fine, Taub, Reynolds, & Kanzler, 2006), and previous research on the WISC-IV found alternative, adequate factor structures that included five factors (Keith et al., 2006; Weiss,

Keith, Zhu, & Chen, 2013); this is in addition to research that found adequate two- and three-factor models (Kaufman, Lichtenberger, & McLean, 2001; L. C. Ward, Ryan, & Axelrod, 2000) for previous versions of the WAIS. The result of this work is that the developers used the five-factor model as the basis for confirmatory factor analysis and found adequate fit using the higher-order model of fit. Canivez and Watkins (in press) found a four-factor model, in which Visual Spatial and Fluid Reasoning were collapsed back into the Perceptual Reasoning factor, a superior fit to the five-factor model. The practical implication is that, first, practitioners should be careful about overinterpreting differences in these two factors. Second, further research will need to be conducted to determine the potential incremental validity and clinical utility of separating out the Visual Spatial and Fluid Reasoning dimensions.

ASSETS AND LIMITATIONS

Since their initial publication, the Wechsler intelligence scales have been used in numerous research studies and are widely used throughout the world. Thus, they are familiar to both researchers and practitioners and have a long and extensive history of continued evaluation. This enormous research base allows practitioners to make relatively accurate predictions regarding clients. Inconsistencies between an individual's performance and relevant research can also be noted, alerting the practitioner that he or she needs to develop and pursue further hypotheses. Furthermore, the subtests are relatively easy to administer, and the accompanying manuals provide clear instructions, concise tables, and excellent norms.

Norms for both the WAIS-IV and WISC-V represent a further clear strength. The size is adequate and, for the most part, has corresponded to the demographics of the U.S. census. Cross-national use has been developed through research on how residents in other countries perform. Sampling on the WAIS-IV and WISC-V for African American and Hispanic individuals closely approximated U.S. census data. A further important feature is that the WAIS-IV was conormed with the WMS-IV and the WIAT. Thus, a high degree of confidence can be placed in comparing scores between these tests. Finally, the WAIS-IV has extended its age range to include the performance for persons in the 70- to 90-year range. This is an important feature, given the increases in knowledge related to this age group along with the expanding number of persons over 65.

Of perhaps even more practical importance to the clinician is the clear, precise data obtained regarding the person's cognitive functioning from the IQ, index, and subtest scores. For example, high scores on the Verbal Comprehension Index indicate good verbal abilities and that the person has benefited from formal education. In contrast, a low score on Processing Speed suggests the person would have difficulty processing information, especially nonverbal information, quickly. Clinicians can become extremely sensitive to the different nuances and implications of various patterns of scores. Thus, many of these interpretive guidelines, particularly for the IQ and index scores, have substantial theoretical and empirical support.

A final, but extremely important, asset of the Wechsler scales is their ability to aid in assessing personality and emotional variables. This assessment can be done by directly

observing the individual as he or she interacts with the examiner, studying the content of test item responses, or evaluating information inferred from the individual's pattern of subtest scores. For example, a person who scored low on Digit Span, Arithmetic, and Coding is likely to be experiencing anxiety, to have an attentional deficit, or a combination of both. In contrast, it might be hypothesized that a person who scored high on Comprehension is likely to have good social judgment. Despite attempts to establish patterns of subtest scores for different psychiatric groups, few clear findings have emerged (Piedmont, Sokolov, & Fleming, 1989a, 1989b). Thus, the Wechsler scales should not be seen as similar to "personality scales" or "clinical scales." Rather, the subject's subtest patterns, behavior surrounding the test, and qualitative responses to the items should be considered as a means of generating hypotheses related to personality. In this context, the Wechsler intelligence scales are noteworthy in the degree to which they can provide personality variables and clinical information.

One significant criticism leveled at the Wechsler scales has been the lack of data supporting their ecological (or everyday) validity (Groth-Marnat & Teal, 2000; Reinecke, Beebe, & Stein, 1999; Sbordone & Long, 1996). Knowing a test's ecological validity is particularly important as referral questions are increasingly related to a client's everyday levels of functioning (e.g., extent of disability, ability to function independently, everyday aspects of memory). Although the Wechsler scales have been correlated with other measures, including the Stanford-Binet and academic achievement, for the most part, there has been a notable lack of comparisons with behavior external to the scales themselves, despite the belief that many significant areas of a person, such as adaptive behavior, personal competence, and need for achievement, are related constructs (Greenspan & Driscoll, 1997; Sternberg, 2003). In particular, the meanings associated with subtest scores should be investigated in more depth. For example, Picture Completion has traditionally been considered a measure of a person's ability to distinguish relevant from irrelevant details in his or her environment, yet this assumption has not been adequately tested. Likewise, no studies have been made to determine if high or low Digit Span scores relate to actual day-by-day behaviors, such as recalling telephone numbers, facility with computer programming sequences, or following directions.

An extension of this concern is that a number of authors have criticized what they believe is an overinterpretation of subtest and index scores (Glutting, Watkins, Konold, & McDermott, 2006; Konold, Glutting, McDermott, Kush, & Watkins, 1999). Specifically, they believe that individual subtest reliabilities are too low and that subtests are not sufficiently specific for interpreting individual outcomes. For example, using the WISC-IV, they noted that, compared with *g* (as represented by the Full Scale IQ), individual index scores did not account for a sufficient proportion of the variance in predicting achievement (Glutting et al., 2006). As a result, index interpretation does not demonstrate sufficient incremental increases in prediction. In addition, the ipsative (within-person) patterns of subtest strengths and weaknesses are not sufficiently stable over time (Macmann & Barnett, 1997). Clinicians might therefore be advised to rely on the Full Scale IQ rather than index scores when making academic (and possibly other) predictions or decisions. Various authors counter this belief by emphasizing the importance of hypothesis testing, combining interpretations with external criteria, and noting the conceptual importance of the complexity of intelligence (Kaufman, 1994,

2000; Kaufman & Lichtenberger, 1999, 2000, 2002, 2006; Lezak, 1988; Lezak et al., 2012; Milberg et al., 1996).

There are several additional limitations to the Wechsler scales. Some critics believe that norms may not be applicable for ethnic minorities or persons from lower socioeconomic backgrounds (see the next section, “Use with Diverse Groups”). In addition, the complexity of scoring, particularly the numerous calculations required for the Wechsler intelligence scales, is likely to increase the probability of clerical errors by examiners (Linger, Ray, Zachar, Underhill, & LoBello, 2007; Loe, Kadlubek, & Marks, 2007; Slate & Hunnicutt, 1988; Slate, Jones, & Murray, 1991). A further potential difficulty is that when supplementary subtests are substituted for core subtests, it is unclear how these supplementary subtests will affect the Full Scale IQ or index scores. As a result, supplementary subtests should be used in calculations only under unusual circumstances, such as when one of the core subtests has been spoiled.

A further issue is that there is a certain degree of subjectivity when scoring many of the items on some subtests. Thus, a “hard” scorer may develop a somewhat lower score than an “easy” scorer. This is particularly true for Similarities, Comprehension, and Vocabulary, where scoring criteria are less clear than for other subtests. The Wechsler scales, like other tests of intelligence, are also limited in the scope of what they can measure. They do not assess some important factors, such as need for achievement, motivation, creativity, or success in dealing with people (Gardner, 2006; Sternberg, 2003).

Finally, it should be noted that the WAIS-IV and WISC-V have continued the traditional measurement of intelligence as represented by the Stanford-Binet scales and the earlier versions of the Wechsler scales. Although their revisions have provided features such as updated norms and index scores (especially the inclusion of Working Memory and Processing Speed as well as Fluid Reasoning on the WISC-V), the underlying theories and essential construction of these scales have remained relatively unchanged for well over 50 years, despite numerous developments in both theory and measurement. These developments include the CHC theory (see Flanagan & Kaufman, 2009), Luria’s PASS (*Planning-Attention-Successive-Sequencing*: Luria, 1980) model, Gardner’s independent competencies (Gardner, 2006), various theories on emotional intelligence (Bar-On, 1998; Ciarrochi, Chan, & Caputi, 2000), and commonsense problem solving (Sternberg et al., 1995). Thus, one criticism of the Wechsler intelligence scales is that they have not responded to more current views on intelligence (Kaufman & Lichtenberger, 2002, 2006; Sternberg, 2003; Sternberg & Kaufman, 1998). However, the fact that the Wechsler scales still manage to make meaningful predictions with other important psychological and life factors provides a basis for their continued use.

USE WITH DIVERSE GROUPS

Each of the considerations discussed in Chapter 2 (see “Assessing Diverse Groups”) should be taken into account when evaluating persons from diverse cultural backgrounds. Incorporating culture into assessment interpretation involves evaluating

a client's level of acculturation and language proficiency, as well as the cultural competency of the examiner. In addition, a degree of targeted, purposeful flexibility, within reason, should occur both when conducting the assessment and when making interpretations. Different accommodations and strategies need to be made based on the outcome of these considerations.

One of the key issues when assessing diverse clients is determining the extent that the instruments used might be biased. Evaluating for test bias can be partially informed by the considerable research devoted to evaluating the extent that intelligence tests such as the Wechsler intelligence scales are biased when used to assess various minority groups in the United States. The majority of these studies have been done with groups that are reasonably well acculturated and with moderate to good English skills. Reviews of this research have generally concluded (e.g., R. M. Kaplan & Saccuzzo, 2005; Sattler, 2008) that intelligence tests are not as biased as has been assumed. For example, deletion of items that appear to have biased content seems to make little difference in overall scores (Sandoval, 1979). Additionally, research has found that the Wechsler scale holds up across different subcultures even in non-Western cultures (Chen, Keith, Weiss, Zhu, & Li, 2010). Numerous validity studies have also found that intelligence tests make academic predictions as accurately for minority groups as for majority groups (e.g., Weiss, Prifitera, & Roid, 1993; Sattler, 2008), and factor-analytic research indicates that the same construct is being measured across various minority groups (Gutkin & Reynolds, 1981). Finally, Japanese populations, who come from a quite different culture from that of the United States, had mean scores that were actually higher than the U.S. standardization group (Lynn, 1977).

Thus, research supports the use of the Wechsler intelligence scales with minority groups in the United States. However, certain groups do indeed perform differently, with advantage generally going to majority groups (Goldbeck, Daseking, Hellwig-Brida, Waldmann, & Petermann, 2010; Holdnack, Drozdick, Iverson, & Weiss, 2013; Lichtenberger & Kaufman, 2012). The central problem does not seem to be the tests themselves, though. Although these tests are far from perfect, they do provide the sort of information they were intended to provide. The main problem seems to be unequal opportunities that are accurately reflected in how various disadvantaged groups perform on intelligence tests. Despite the conclusion that cognitive tests generally measure what they intend to measure, clinicians still need to take extra care to ensure that accurate data and conclusions are developed. When testing individuals from different cultural backgrounds, these general and specific guidelines seem appropriate:

- Make extra efforts to ensure that clients feel comfortable and welcomed.
- Make extra efforts to increase motivation; encourage clients to do their best.
- Make sure that communication is as clear as possible, especially if there are differences in accents between the clinician and the client.
- Resources beyond merely tests should have a greater significance with diverse clients (e.g., teacher reports, discussions with parents, history, behavioral observations) than for majority clients.
- If language and culture appear to have been a factor in lowering the client's performance, subtests that seem to be less influenced by culture and language

should be the focus of interpretation. (Deemphasize language-based subtests, such as Vocabulary and Information, and instead emphasize nonverbal tests, such as Block Design, Matrix Reasoning, and Visual Puzzles.)

- When assessing persons from cultures that deemphasize performing tasks rapidly (e.g., South Pacific islands), deemphasize speeded subtests (Processing Speed Index; Coding, Symbol Search, Cancellation).
- Be cautious interpreting $PRI < VCI$ differences for African Americans and Native Americans with right-hemisphere damage; the differences have been found to be less meaningful when compared with European Americans (based on $PIQ < VIQ$ for the WAIS-R and WAIS-III; see Kaufman & Lichtenberger, 2006).
- Be cautious interpreting $VCI < PRI$ differences for African Americans with left-hemisphere damage; the expected verbal-nonverbal difference found among European Americans has not been found in this population (based on $VIQ < PRI$ for the WAIS-R and WAIS-III; see Kaufman & Lichtenberger, 2006).

When clinicians determine that clients are highly identified with their culture and have minimal proficiency with English, many of the same considerations still apply. First, though, clinicians should find out if there is a version of the test that has been adapted for the culture and primary/preferred language of the client (e.g., Wechsler, 2008a). Additionally, clinicians should make extra efforts to ensure that clients are comfortable, made to feel welcome, and encouraged to do their best and that communication is clear. Nontest information should also be considered carefully. If it is decided to administer the Wechsler intelligence scales, nonverbal subtests should be given greater emphasis. Verbal-oriented tests should never be used for interpretations for clients who do not have at least adequate English language proficiency. In some cases it might be advisable to use an interpreter, if there is no option in the client's primary language or if there is no access to an assessor who could administer it in that language. If an interpreter is used, it would be important to locate one who is familiar not only with the client's language but with the person's values, culture, and ideology. However, using an interpreter also means that the test administration will not be standard; as a result, there may be a reduction in test validity. In particular, it may be more difficult to translate some of the directions and responses. The meaning and level of difficulty of some of the items might change. For example, vocabulary items might be either more or less difficult in the client's native language. Thus, it may be advisable to use one of the 20 language translations of the Wechsler intelligence scales.

One strategy for assessing clients who are highly identified with their own culture and do not have proficiency with English may be to use alternative nonverbal tests. These might be used in conjunction with the Wechsler intelligence scales or be the sole means of assessing intelligence. Examples are the Comprehensive Test of Nonverbal Intelligence-2 (CTONI-2; Hammill, Pearson, & Wiederholt, 2009); Universal Nonverbal Intelligence Test-2 (UNIT-2; Bracken & McCallum, 2015); the Test of Nonverbal Intelligence-4 (TONI-4; Brown, Sherbenou, & Johnsen, 2010); and the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006). One of the main motivations for developing these tests was for use with quite diverse populations who do not have English proficiency. However, even some groups with good English

proficiency might benefit from more nonverbally oriented procedures. For example, Native Americans have been found to score up to 25 or 30 points higher on nonverbal tests when compared with verbal tests (McShane & Plas, 1984). Thus, using nonverbal tests will illustrate their strengths. In addition, measures of a client's ongoing learning abilities (so-called dynamic testing) show promise in assessing the extent to which a client can benefit from various learning environments (learning potential; Sternberg & Grigorenko, 2001).

Persons with visual, hearing, or motor disabilities present different challenges for clinicians. If clients have visual impairments, subtests with visual components cannot be administered (e.g., Coding, Symbol Search, Block Design, Matrix Reasoning). Instead, greater reliance will need to be placed on auditory/verbal subtests. Different considerations need to be made for persons with hearing impairments. For these clients, more reliance will need to be given to nonverbal tests. Specialty nonverbally oriented tests might be considered. Sometimes clinicians may decide to use an interpreter (American Sign Language [ASL]). While this means that a greater variety of tests can be administered, it also means the administration is nonstandardized, with the resulting potential for reduced validity. Clinicians will need to evaluate whether the greater variety of tests that can be administered will compensate for the potential loss of validity. The *WAIS-IV Administration and Scoring Manual* provides recommendations on which subtests to use with different types of interpreters (e.g., ASL versus Cued Speech). For example, the Arithmetic subtest does not seem to be significantly altered with ASL, but, because sign language supplies unintended cues, the Vocabulary subtest does seem to be significantly modified and therefore may not produce valid scores.

MEANING OF IQ SCORES

Because only a weak and vague relation exists between theories of intelligence and the Wechsler intelligence scales, it is important for all persons involved with testing to understand the meaning of IQ scores. Untrained persons are particularly likely to misinterpret IQ scores, which may result in poor decisions or negative attitudes toward the client, the clinician, or the testing procedure itself. The meaning of IQ scores can be partially clarified by elaborating on some of the more common misinterpretations. IQ is often incorrectly believed to be fixed, unchangeable, and innate. Although there does tend to be considerable stability of IQ scores throughout adulthood ($r = .85$; Schuerger & Witt, 1989), it is possible for changes in IQ to occur, particularly among children (Perkins & Grotzer, 1997). For example, the greatest longitudinal increases in IQs occurred among children who were from homes that provided strong encouragement and avoided severe forms of punishment (McCall, Appelbaum, & Hogarty, 1973). Similarly, Sameroff, Seifer, Baldwin, and Baldwin (1993) found that multiple environmental risk factors (e.g., number of major stressful events, mother's mental health) were able to predict one-third to one-half of IQ variance between the ages of 4 and 13. In addition, education can increase aspects of IQ primarily related to crystallized intelligence even among adults. Thus, IQ can be related to a number of environmental influences.

In addition, IQ scores are not exact, precise measurements; rather, they are estimates in which there is an expected range of fluctuation between one performance and the next. Finally, tests such as the Wechsler scales measure only a limited range of abilities, and a large number of variables usually considered “intelligent” are beyond the scope of most intelligence tests. No test or battery of tests can ever give a complete picture; tests can only assess various areas of functioning. In summary, an IQ is an estimate of a person’s current level of functioning as measured by the various tasks required in a test.

An assumption of any global IQ score is that it derives from a wide array of interacting abilities. A subtest such as Information assesses specific areas of a person’s range of knowledge and is related to general intelligence. However, optimal performance on the Information subtest is influenced by achievement orientation, curiosity, culture, the person’s interests, and educational opportunities. More general prerequisites are that the client must comprehend what has been requested, be motivated to do well, follow directions, provide a response, and understand English. Factors such as persistence and drive are also likely to influence any type of task presented to the person. The tasks included in IQ tests are those, based on judgments by psychometrists, most valued by Western society. In other words, they relate to and are predictive of relevant skills outside the testing situation. It is certainly possible to test a much wider range of areas (as in Guilford’s *Structure of Intelligence*), but this is not routinely done since those areas often have little relevance in predicting academic achievement or vocational performance.

Despite the many relevant areas measured by IQ tests, practitioners need to observe some humility when making predictions based on them. Many persons with quite high IQs achieve little or nothing. Having a high IQ is in no way a guarantee of success; it merely means that one important prerequisite has been met. In contrast, persons with relatively low IQs have more severe limitations placed on them. As a result of their relatively narrower range of options, predictions regarding their behavior tend to be more accurate. However, it is possible that persons with average or below-average WAIS-IV/WISC-V IQs may have high levels of interpersonal, practical, or emotional “intelligence,” which may help them compensate for lower levels of formal intelligence.

Regardless of the person’s IQ range, clinicians should be clear regarding the likely band of error (SEM). It is often useful to include the SEM in a report. For example, the WAIS-IV Full Scale IQ has an average SEM of 2.16 (Wechsler, 2008b). Thus, a particular IQ has a 95% chance of being within 2.16 IQ points of a person’s obtained IQ. The WISC-V has a slightly higher average SEM of 2.90 for the Full Scale IQ (Wechsler, 2014b). Error can also be the result of unforeseen events beyond the context of IQ tests. Although 50% to 75% of the variance of children’s academic success is dependent on nonintellectual factors (e.g., persistence, personal adjustment, family support), most of a typical assessment is spent evaluating IQ. Some of these nonintellectual areas might be quite difficult to assess, and others might even be impossible to account for. For example, a student might unexpectedly develop an excellent relationship with a new teacher, which significantly changes his or her attitude toward school, thereby stimulating his or her interest to passionately pursue a specific area. Thus, any meaning attached to an IQ score should acknowledge the possible effects of uncertainty both in the measurement itself and from the wider context of the person’s life.

Another important aspect of IQ is the statistical meaning of the different scores. Binet originally conceptualized intelligence as the difference between a person's mental age and his or her chronological age. Binet's formulation was found to be inadequate and has been replaced by the deviation IQ. The assumption behind the deviation IQ is that intelligence falls around a normal distribution (see Figure 5.1). The interpretation of an IQ score, then, is straightforward because it gives the relative position of a person compared with his or her age-related peers. The IQ can thus be expressed in deviation units away from the norm. The Wechsler Full Scale IQ and the four/five indexes have a mean of 100 and a standard deviation of 15. Scores also can be easily translated into percentile equivalents. For example, an IQ of 120 is 1.33 standard deviations above the mean and places an individual in the 91st percentile (see *WAIS-IV Administration and Scoring Manual*, Tables A.3 to A.7, and *WISC-V Administration and Scoring Manual*, Table A.7, for conversions). Thus, this person's performance fell at a score higher than 91% of his or her age-related peers. The IQ cutoff for mental retardation is around 70, which indicates that such individuals are functioning in the lowest 2% when compared with their age-related peers.

A final consideration is the different classifications of intelligence. Table 5.5 lists commonly used diagnostic labels and compares them with IQ ranges and percent-ages. These terms are taken from the 2008 WAIS-IV Record forms. Thus, an IQ can be expressed conceptually as an estimate of a person's current level of ability, statistically

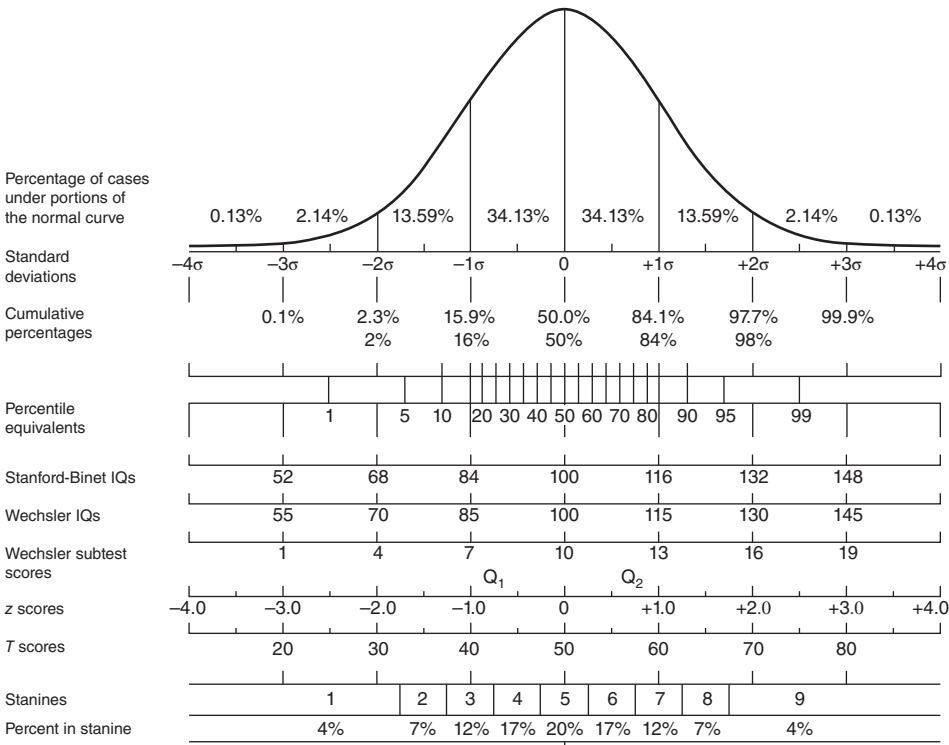


Figure 5.1 Relationship of Wechsler scores to various types of standard measures

Table 5.5 WAIS-IV/WISC-V Intelligence Classifications

Classifications	More value-neutral terms	Corresponding IQ range
Very superior	Upper extreme	130+
Superior	Well above average	120–129
High average	High average	110–119
Average	Average	90–109
Low average	Low average	80–89
Borderline	Well below average	70–79
Extremely low	Lower extreme	69 and below

Source: Intelligence classifications adapted from WAIS-IV and WISC-V Record forms.

as a deviation score that can be transformed into percentile equivalents, and diagnostically using common terms for classification.

CAUTIONS AND GUIDELINES IN ADMINISTRATION

The Wechsler manuals generally provide quite clear guidelines for administration and scoring. Despite this clarity, the number of administration and scoring errors on the part of both trainee and experienced clinicians is far higher than it should be (Alfonso, Johnson, Patinella, & Radar, 1998; Linger et al., 2007; Loe et al., 2007; Moon, Blakey, Gorsuch, & Fantuzzo, 1991; Moon, Fantuzzo, & Gorsuch, 1986; Slate & Hunnicutt, 1988; Slate et al., 1991). These errors can affect scores dramatically, which can in turn affect conclusions and recommendations. One way of reducing clerical errors is to use the computer scoring software developed by Pearson Assessment (i.e., WAIS-IV Scoring Assistant, WAIS-IV Writer, WISC-V Scoring Assistant, WISC-V Writer). Even with repeated administration of the Wechsler scales, examiners often end up practicing their mistakes rather than correcting them (Slate et al., 1991). The causes of these errors include lack of proper instruction, lack of clarity between academic versus clinical site regarding where training is supposed to occur, carelessness, variations in the quality of the examiner–examinee relationship, and work overload for clinicians (Slate & Hunnicutt, 1988). One approach to reducing errors is awareness regarding the most frequent general categories of errors. The most common errors have been found to be (Loe et al., 2007; Slate & Hunnicutt, 1988):

1. Failing to query verbal responses.
2. Assigning too many points to an answer (leniency by examiner).
3. Failing to record examinee responses, circle scores, or record times (errors of administration).
4. Failing to question responses when required by test manual (poor reading and recalling of information in the manual).
5. Questioning examinee inappropriately (poor reading and/or incorrect integration of the manual).

6. Assigning too few points when required by test manual (examiner scores too hard).
7. Incorrectly converting raw score to standard score (clerical error).
8. Failing to assign correct points for nonverbal items (clerical and timing errors).
9. Incorrectly calculating raw scores for subtest totals (clerical error).
10. Incorrectly calculating chronological age (clerical error).

Whereas this list covers quite general categories, Moon et al. (1991) have developed a list of the most frequently occurring specific errors along with concrete, specific recommendations to correct for these errors. However, it should be cautioned that these errors may not necessarily be the ones that occur most frequently for the latest versions of the Wechsler intelligence scales (WAIS-IV/WISC-V):

1. Recite digits (on Digit Span) and digits and letters (on Letter-Number Sequencing) at the rate of 1 per second with the pitch of the voice dropping on the last digit/letter of each trial.
2. State during the introduction that each task begins with easy questions and ends with difficult ones. Examiners may also note that not everyone is expected to succeed on all problems.
3. Record responses verbatim on Vocabulary. At times, the examinee provides so much detail that this is not possible, but the essential components should be written down verbatim. The use of abbreviations can facilitate such recording.
4. Properly orient blocks (on Block Design) at the examinee's midline.
5. The first time the examinee points out a nonessential part on Picture Completion, the examiner should say, "Yes, but what is the most important thing missing?"
6. Attempt to elicit the examinee's perception of the testing situation and correct any misconceptions.
7. Check to see if the examinee is comfortable.

Despite clear guidelines in the manual as well as awareness of frequent errors, examiners are still likely to make mistakes. Thus, optimal training guidelines should be incorporated into graduate programs and continuing education. A recommended format is the Mastery Model, which involves five steps:

1. 1 to 2 hours studying the manual.
2. Viewing a videotape of an error-free WAIS-IV/WISC-V administration.
3. Viewing a videotaped lecture of major pitfalls of administration.
4. Successfully detecting errors in a videotaped flawed WAIS-IV/WISC-V administration.
5. Actually administering the WAIS-IV/WISC-V to be evaluated by a rating device such as Sattler's (2008) "General Test Administration Practices Checklist." Such procedures are likely to significantly shorten the length of training time and number of training administrations and yet significantly increase the level of competence related to Wechsler scale administration and scoring (Moon, Fantuzzo, & Gorsuch, 1986; Slate et al., 1991).

The *WAIS-IV Administration and Scoring Manual* indicates that the average time to administer the 10 core subtests used to determine the Full Scale IQ and four index scores for the normative sample ranged between 67 and 100 minutes. Including all the supplemental subtests took an average additional 20 to 24 minutes. However, research with the WAIS-III found that, for a heterogeneous clinical population, the average time was somewhat longer than that reported in the manual (Ryan, Lopez, & Werth, 1998). It may be that administration time for the WAIS-IV among clinical populations is longer than the time estimated in the manual. Time estimates given in the WISC-V manual to administer the 10 core subtests used to calculate the Full Scale IQ and five index scores ranged from 56 to 70 minutes. Research on the WISC-IV with a school population found that the range was a bit wider (42–100+ minutes) with an average length of 72 minutes (Ryan, Glass, & Brown, 2007). These times were for administration only and did not include time required for scoring, breaks, or interpretation, which usually takes an additional 50 to 60 minutes. One practical implication is that clinicians may need to take extra time for some clients, especially those who are slow, fatigue easily, or provide overly detailed responses. In some cases it might be necessary to assess such clients over two separate sessions. Finally, clinicians should make realistic appraisals of required times and use these estimates to make sure that they are appropriately compensated.

WAIS-IV/WISC-V SUCCESSIVE-LEVEL INTERPRETATION PROCEDURE

The successive-level approach to interpreting Wechsler scores represents an integration and synthesis of the approaches outlined by major resources in the field (Flanagan & Kaufman, 2009; Kaufman & Lichtenberger, 2006; Sattler, 2008; Weiss, Saklofske, Holdnack, & Prifitera, 2015). This approach provides clinicians with a sequential, five-level format for working with and discussing a person's performance. The underlying purpose for each of these steps should be based on confirming, disconfirming, or altering hypotheses based on patterns of scores combined with relevant background information. The next section of this chapter ("Wechsler Indexes and Subtests") covers descriptions of the Wechsler subtests, including the more frequently encountered abilities associated with these subtests. This section can serve as a summary and quick reference for clinicians, especially in analyzing test profiles (Levels II and III).

Examiners who are relatively unfamiliar with the Wechsler scales are likely to find the level of detail in these interpretation procedures and Wechsler subtest sections somewhat daunting because of their complexity. It is recommended that they read the interpretation procedures first to gain familiarity with the material. It might be particularly helpful to review the summary of these procedures in Table 5.6 both before and after reading this section. The table can also serve as a useful future quick reference guide when actually working with Wechsler protocols. After perusing the "Wechsler Indexes and Subtests" section, student examiners should next obtain a completed WAIS-IV/WISC-V profile, preferably one they themselves have administered, and then work through the levels of interpretation in a sequential manner. Doing this should add the required level of clarity and integration of the material to enable examiners to work more confidently with future protocols.

Table 5.6 Summary of Successive Five-Level WAIS-IV/WISC-V Interpretive Procedures**Level I.** Interpret the Full Scale IQ

Determine percentile rankings and IQ classifications

Level II. Interpret index scores and CHC groupings

Interpret *personal* strengths and weaknesses (*ipsatively*) for the examinee if significant discrepancies occur between clusters of index scores; *normative* interpretations can still be made whether significant differences occur or not

- a. *Index scores*: Verbal Comprehension, Perceptual Reasoning (Visual Spatial and Fluid Reasoning on the WISC-V), Working Memory, Processing Speed
- b. *CHC/other cluster groupings*: Fluid Reasoning, Verbal Fluid Reasoning, Nonverbal Fluid Reasoning, Quantitative Reasoning, Lexical Knowledge, General Information, Visual Processing, Cognitive Proficiency, Visual Motor Speed, Problem Solving without Visual Motor Speed, Long Term Memory, Auditory Working Memory/Short Term Memory (note that all core and supplemental subtests must be given to calculate CHC groupings)

Level III. Interpret subtest variability**Level IV.** Qualitative/process analysis**Level V.** Analyze intrasubtest variability

An ideal set of interpretive statements is not only accurate but also gives the sense of what the cognitive ability measures and connects it with the client's everyday functioning. It should ideally use everyday language with a minimum of technical terms. However, this is often a difficult interpretive skill to develop. The following five general strategies might be used to expand on cognitive interpretations:

1. Make an initial general statement ("excellent verbal abilities [VCI = 125, top 5% of age-related peers]," etc.).
2. Elaborate by listing subcomponents of ability ("good fund of general information, word knowledge.").
3. Give qualitative description of test responses ("could easily define quite difficult words" or "accurately described the similarities between two related objects or ideas") or test items ("was able to recall the accomplishments of famous people and accurately describe a variety of scientific facts").
4. Give qualitative description of history/behavioral observations ("She quickly, concisely, and accurately answered questions.").
5. Provide implications for everyday life ("She would be able to easily understand complex conversations" or "would do well in verbally oriented occupations").

Rarely is it necessary to use all of these strategies. In some cases it might be necessary to use only the first one. In other situations, especially when the interpretation is crucial to the referral question, it might be important to use three or four of them. Two cautions related to the third strategy: first, the actual items should not be described, since this would be a breach of test security; instead, a general description or alternative examples illustrating the types of items that were administered and answered might

be provided. Second, including these descriptions should not obscure the findings or confuse or overwhelm the reader.

These are principles to keep in mind when working through the interpretation procedures:

- The successive steps begin with the most general aspects of the WAIS-IV/WISC-V (Full Scale IQ) and gradually work their way to more specific aspects of the person's performance (indexes, CHC groupings, qualitative responses to individual items, etc.).
- Examiners can interpret the more global measures (Full Scale IQ, Global Ability Index) with greater meaning, usefulness, and certainty if there is not a high degree of difference among the index scores (23 points for both the WAIS-IV and WISC-V between the highest and lowest index scores) or other groupings. With increasing differences, the purity of the global measures becomes contaminated so that interpretations of them become less meaningful. For example, if there is a pattern in which the Verbal Comprehension and Perceptual Reasoning Indexes are significantly different from each other, it makes more sense to focus on these two indexes rather than the more global Full Scale IQ (or Global Ability Index).
- The recommended level set to establish significant difference is the .05 level. This is true for differences through all levels of interpretation, including indexes and additional groupings. It is felt that this level of significance is sufficiently rigorous for clinical purposes. If either less stringent ($p = .15$) or more stringent ($p = .01$) levels are desired, relevant tables can be found in the WAIS-IV and WISC-V *Administration and Scoring Manuals* (Wechsler, 2008a, 2014a). When possible, Bonferroni corrections have been included to correct for the possible statistical error resulting from inflation of significant results because of the number of comparisons.
- To determine whether index scores are significantly (.05 level) discrepant from each other, tables can be consulted in the manuals (*WAIS-IV Administration and Scoring Manual*, Table B.1, p. 230; *WISC-V Administration and Scoring Manual*, Table B.1, p. 342). Thus, comparisons are made between the different pairs of indexes.
- In contrast to the previous procedure for discrepancies between indexes, subtest fluctuations are based on *comparisons with mean scores*. One strategy is to compare the scaled score of each individual subtest with the mean for all the subtests administered (and then calculate the difference that the subtests fluctuate from this mean to see if the difference is significant). An alternative strategy is to find the mean of the subtests within the index scores (rather than for the mean for the 10 core subtests) to see if the subtest scores are significantly different.
- Any interpretations, especially those related to the more specific levels (Levels III, IV, and V), should be considered as tentative hypotheses requiring support from additional sources of information (behavioral observations, school records, etc.). Preferably, each hypothesis should be supported by at least two additional sources. This process of hypothesis generation, confirmation/disconfirmation, and integration with other sources is not merely a statistical procedure but also involves considerable clinical wisdom and judgment.

Level I. Full Scale IQ (and Global Ability Index)

An examinee's Full Scale IQ should be considered first because it provides the basis and context for evaluating other cognitive abilities. It is generally the single most reliable and valid score. The Full Scale IQ gives the person's relative standing in comparison with his or her age-related peers and provides a global estimate of his or her overall mental abilities. It is often useful to transform the Full Scale IQ into a percentile rank (see *WAIS-IV Administration and Scoring Manual*, Tables A.3–A.7, pp. 220–225; *WISC-V Administration and Scoring Manual*, Appendix A.7, pp. 334–335) or intelligence classification (see Table 5.5 or WAIS-IV/WISC-V record forms). This is especially important when relating test results to untrained persons because both percentile rank and intelligence classifications are usually less subject to misinterpretation than IQ scores. Many examiners also prefer to include the SEM as an estimate of the confidence (“confidence intervals”) that can be placed in the obtained score. For example, an achieved Full Scale IQ of 110 means that there is a 95% probability that the individual's true IQ falls between 105 and 115. This clarifies that the IQ score is not a precise number but is rather a range with an expected margin of error. However, including confidence intervals in actual reports is likely to be more technical detail than most referral sources need. Additionally, one of the classifications, “Borderline,” might potentially be misinterpreted, because it might be confused with the *DSM-5* psychiatric diagnosis of Borderline Personality Disorder. Examiners might counter this by clarifying in parentheses that the “Borderline” range can also be described as “Well below Average.”

Although the Full Scale IQ is the most stable and well-validated part of the Wechsler scales, its significance becomes progressively less important as the fluctuations increase between index scores and other groupings. That is, when there are significant differences among the indexes, the Full Scale IQ becomes less easily interpretable and can in fact be misleading. When such fluctuations occur, it is incumbent on the examiner to work in more detail to extract the significance of these relative strengths and weaknesses. The next four successive levels of interpretation provide a sequential method of accomplishing this goal.

The Global Ability Index (GAI) is an alternative global measure of intelligence. On the WAIS-IV, it is calculated by using only the three subtests used for the Verbal Comprehension Index and the three subtests used for the Perceptual Reasoning Index. On the WISC-V, it is calculated using two verbal subtests, one visual spatial subtest, and two fluid reasoning subtests. Therefore, on both, it eliminates using any Working Memory and Processing Speed subtests. This is important because Working Memory and Processing Speed are the indexes that are most sensitive to deterioration and most susceptible to situational factors (e.g., lack of sleep, hunger, depression, etc.). This finding is consistent with the finding that 65% of examinees with a wide variety of cognitive difficulties had GAI index scores that were significantly higher than their Full Scale IQs (Wechsler, 2008b). Thus, calculating the difference between Full Scale IQ and GAI provides an index of the degree to which brain-, age-, and situation-sensitive subtests are lowering a person's overall level of functioning (see *WAIS-IV Technical and Interpretive Manual* for guidelines). The general descriptions noted for describing the Full Scale IQ also apply to the GAI.

Level II. Indexes and Additional Groupings

The second level of interpretation is to consider index scores and additional groupings. The .05 level of significance is consistently used to determine if fluctuations are significant. In some cases, procedures and formulas are provided to determine the significance of various fluctuations and to convert scores into the familiar IQ-related standard scores having a mean of 100 and standard deviation of 15.

Step IIa. Index Scores

The core level of interpretation is the index scores. These scores represent an optimal level of specificity. In contrast, the Full Scale IQ is so general that it does not provide much information related to a client's strengths and weaknesses. The subtests are quite narrow measures of abilities, but their reliability and validity are sufficiently low that relying on them for interpretation has been seriously questioned (Glutting, Watkins, Konold, & McDermott, 2006; Konold, Glutting, McDermott, Kush, & Watkins, 1999). Thus the indexes provide a sound, empirically and conceptually based means of understanding more detailed aspects of the person's intellectual functioning. Examples are provided to illustrate how high or low scores might translate into everyday functioning.

Indexes should be interpreted only if they represent unitary abilities. A unitary ability is one in which there is less than a 5-point subscale difference among the subtests within the index. If the difference between the highest subtest and the lowest one within the index is 5 or more subscale points, then do not interpret that index. Instead, proceed to Step IIb to investigate whether there are additional meaningful clusters.

One important consideration is whether relative index weaknesses are cause for concern. This depends on how low the person performed relative to the rest of the population. For example, a person may have a Verbal Comprehension Index of 125 but a Processing Speed Index of 100. These scores certainly represent considerable variation such that it is safe to conclude that the speed by which the person processes information is a relative weakness. This can be considered a *personal* (or ipsative) weakness since it is a weakness relative to the examinee's other score. However, since Processing Speed is still in the average range compared to others the person's age, it is not likely to create problems in adapting to most everyday situations. The situation is quite different if there is a similar 25-point Verbal Comprehension–Processing Speed difference but with Verbal Comprehension being 100 and Processing Speed being 75. The Processing Speed score of 75 strongly suggests that this would create difficulties in the person's ability to function adequately if given even minimally demanding tasks, such as clerical procedures. It can thus be considered not only a *personal* (ipsative) weakness, but also a *normative* weakness, since it is quite low compared with the person's age-related peers. In the examples given later of everyday functioning listed under each of the indexes, it is essential to understand that these examples refer to normative strengths/weaknesses rather than merely relative/ipsative strengths/weaknesses.

A further strategy for interpreting index scores is to note whether there are significant differences among various pairs. Discrepancy comparisons are part of the standard scoring procedure for the WAIS-IV and WISC-V. Calculations include whether

the differences are significant and the frequency that the difference occurs in the standardization sample (base rate; see *WAIS-IV Administration and Scoring Manual*, Tables B.1 and B.2, pp. 230–231; *WISC-V Administration and Scoring Manual*, Tables B.1 and B.2, pp. 342–346). For example, if a person's WAIS-IV Perceptual Reasoning Index was 15 points higher than his or her Processing Speed Index ($PRI > PSI$), this difference would clearly exceed the .05 level for all age groups (as per Table B.1), and, according to Table B.2, a $PRI-PSI$ difference of 15 points occurred in only 15% of the standardization sample. The next and more challenging step is to determine what the significance is for this difference. On one hand, it might simply represent natural variations in the person's abilities. Indeed, it is quite natural for variations in abilities to occur for many people, especially for persons in the high-IQ range. In the example above, of the relatively higher Perceptual Reasoning versus Processing Speed, this difference might reflect a person who worked in a nonverbal area (e.g., skilled craftsman) and had a corresponding slow, deliberate work style. Thus, it may not be clinically significant. In contrast, another person may have the same pattern but may have sustained a recent head injury. The slow Processing Speed might reflect the impact of this injury. This would especially be the case if the person were employed in an occupation that required rapid processing of information. The much slower speed would then have practical significance in that the person might have a particularly difficult time readjusting to work.

Step IIb. Additional Clusters

Interpreting additional clusters of subtests involves various subtests organized according to CHC and WAIS-IV/WISC-V theoretical concepts (Flanagan & Kaufman, 2009; Flanagan & Ortiz, 2001; Lichtenberger & Kaufman, 2012). As with the index scores, it is first necessary to decide whether the clusters represent unitary abilities. A unitary ability within a cluster is defined by there being fewer than 5 scaled score points among the subtests comprising the cluster. Five or more points difference within the subtests of the cluster suggest that the cluster score is not unitary. Accordingly, the cluster should not be interpreted. It is then incumbent on the clinician to determine whether there are other combinations of subtests that help explain the client's cognitive abilities. Lichtenberger and Kaufman (2012) have identified nine CHC/WAIS-IV subtest clusters for the WAIS-IV (descriptions derived from Lichtenberger and Kaufman, 2012):

1. *Fluid Reasoning* (Matrix Reasoning + Figure Weights, which is an index on the WISC-V). Mental operations required for a novel task where the operations cannot be done automatically; examples include accurately perceiving relationships among patterns, drawing inferences, recognizing and forming concepts, solving problems, making inferences, understanding implications, reorganizing and transforming information, and extrapolating. This cluster can be calculated on the WAIS-IV only for persons between the ages of 16 and 69 since there are no norms for persons above the age of 70 on Figure Weights.
2. *Visual Processing* (Block Design + Visual Puzzles, labeled the Visual Spatial Index on the WISC-V). Perception, generation, synthesis, manipulation, transformation, storage, and retrieval of visual information; examples include

perceiving and manipulating spatial patterns, maintaining spatial orientation, interpreting how objects change as they move through space, mentally reversing and rotating objects.

3. *Verbal Fluid Reasoning* (Similarities + Comprehension). Involves fluid reasoning (see #1) specific to verbal material, includes depth and breadth of acquired knowledge through interacting with culture and then ability to abstract and apply this knowledge.
4. *Lexical Knowledge* (Vocabulary + Similarities). Breadth and depth of accumulated knowledge of a culture and how to apply this knowledge, application of word knowledge, extent to which words are understood and used.
5. *General Information* (Comprehension + Information). Range of general information.
6. *Long-Term Memory* (Vocabulary + Information). Range of general information combined with word knowledge, both of which are stored in long-term memory.
7. *Short-Term Memory* (Letter-Number Sequencing + Digit Span). Holding and using information in immediate awareness, limited to 7 “chunks” of information (plus or minus 2); examples are remembering telephone numbers, ability to hold directions in memory long enough to complete a task. Note that this cluster is also referred to as *mental manipulation* when using terminology from the WAIS-IV theoretical concepts. Its ability/construct is similar to, if not the same as, working memory.
8. *Visual-Motor Speed* (Block Design + Coding + Symbol Search). Rapid information processing for visual information; involves fluid reasoning specific to nonverbal information (in contrast to verbal fluid reasoning).
9. *Problem Solving without Visual-Motor Speed* (Matrix Reasoning + Visual Puzzles + Picture Completion + Figure Weights). Pure measure of nonverbal problem solving since none of the subtests listed requires either coordination or speed of processing. This cluster can be calculated only for persons between the ages of 16 and 69 since there are no norms for persons above the age of 70 for Figure Weights.

The WISC-V includes some clusters within the test itself, and Flanagan and Kaufman (2009) developed clusters for the WISC-IV that can be applied to the WISC-V. A combination of these two sets of clusters is listed next. The clusters are similar to the WAIS-IV clusters, but with slightly different subtests and configurations. For most, the definitions are the same as above. However, there are slight differences:

1. *Nonverbal Fluid Reasoning* (Matrix Reasoning + Picture Concepts). Mental operations required for a novel nonverbal task where the operations cannot be done automatically; examples include accurately perceiving nonverbal relationships among patterns, recognizing and forming nonverbal concepts, solving nonverbal problems, making inferences, reorganizing and transforming nonverbal information, extrapolating.
2. *Quantitative Reasoning* (Figure Weights + Arithmetic). Mental operations related to mathematical and quantitative ability.

3. *General Information* (Comprehension + Information).
4. *Auditory Working Memory* or *Short-Term Memory* (Digit Span + Letter-Number Sequencing). Working memory ability specific to auditory, verbal information; ability to hold auditory and verbal information in mind, manipulate it, and produce some reasoning outcome; holding and using information in immediate awareness, limited to 7 “chunks” of information (plus or minus 2); examples are remembering telephone numbers, ability to hold directions in memory long enough to complete a task. Note that this cluster is also referred to as *mental manipulation*.
5. *Long-Term Memory* (Information + Vocabulary).
6. *Cognitive Proficiency* (Digit Span + Picture Span + Coding + Symbol Search). Speed and accuracy of mental operations; efficiency of mental abilities on novel and difficult tasks.

The cluster scaled scores can be used to make *normative* comparisons. Thus, a Visual Processing score of 85 would indicate the client is 1 standard deviation below his or her age-related peers in this ability. The client would most likely have a difficult time making sense of and solving problems related to what he or she sees. Interventions might include breaking visual information into basic, step-by-step instructions or highlighting the most crucial visual information for the client to notice.

In addition to these normative comparisons, interpretation can also involve understanding the person’s strengths and weaknesses compared with his or her own performance (*ipsative* comparisons). Ipsative comparisons can be made by comparing relevant pairs of clusters. In order for the differences to be interpretable (significantly different), they must vary at or greater than a certain amount, usually around 20 points. For example, if an adult’s Fluid Reasoning is 20 or more scaled score points higher than his or her Visual Processing, it would suggest that a relative strength would be in the individual’s ability to solve novel problems. In contrast, the individual would have much more difficulty with straightforward tasks using visuospatial skills.

Level III. Interpreting Subtest Variability

The next step is to consider the degree to which the individual subtests deviate from the Full Scale IQ or the index scores. This information can then be used to develop interpretations that have been made based on the Full Scale IQ and index scores. The outcome should be a description of a person’s relative cognitive strengths and weaknesses. A listing and discussion of the meaning of each subtest and the abilities it measures is provided in the section of this chapter titled “Wechsler Indexes and Subtests.”

In developing their own hypotheses about important dimensions of high and low scores, clinicians can refer to this section as well as to information on how to assess diverse populations. Readers may also wish to refer to Flanagan and Kaufman (2009), Kaufman and Lichtenberger (2006), Lichtenberger and Kaufman (2012), and Sattler (2008), who have provided detailed lists of hypotheses and useful tables for various combinations of high and low subtest scores. However, Level III interpretation is

necessary only if there is sufficient subtest scatter. If all the subtests are fairly even, it is not necessary to attempt subtest profile interpretation.

As was noted previously, subtest interpretation has been the source of controversy in that some authors have pointed out that the subtests are not sufficiently reliable, do not have enough subtest specificity, and do not provide sufficient incremental validity beyond what might be accounted for by the Full Scale IQ (Konold et al., 1999; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; M. W. Watkins, Glutting, & Lei, 2007). Criticism of subtest interpretation has mainly been based on empirical concerns, but there are also underlying conceptual differences centered on whether intelligence is mainly accounted for by *g* (“lumpers”) as opposed to it being composed of a number of different components (“splitters”). This debate seems to have been present almost as long as conceptions of intelligence have been in existence. One common response to this issue is that subtest interpretation is not merely an empirical activity but also involves a clinical process of hypothesis testing and integrating a variety of sources of data (Kaufman & Lichtenberger, 2006; Lezak et al., 2012).

As a result of the difficulties with subtest interpretation, the major focus should be on the index and CHC/cluster groupings. These provide a much stronger basis for making interpretations. In some cases, however, investigating subtests can be useful. These three steps are recommended to interpret subtest variability:

1. Determine whether subtest fluctuations are significant (see “Determining Strengths and Weaknesses” sections of the Record Form).
2. Develop hypotheses related to the meaning of the relative high/low scores.
3. Integrate these hypotheses with additional relevant information regarding the examinee.

Clinicians should *never* interpret subtests merely by noting what seem to be high/low subtests and then listing the abilities provided in the subtest descriptions. Interpreters who merely list the subtest’s abilities as they are listed in a book will make incorrect and even potentially damaging conclusions about the examinee. Clinicians need to be aware that interpreting subtest variability involves clinical judgment guided by theory, observation, and an integration of the specifics of each case. Because there is little research base to support this process, interpreting subtest variability should be approached with caution. Again, the preferable method of working with subtest variability is to note various clusters (see interpretive level II). Inferences based on one or two subtests should be considered quite tentative.

The challenge with interpreting high/low subtests is that any subtest will involve a series of different abilities. Just because a subtest or group of subtests has been designated as a relative strength or weakness does not mean that it is clear which of the various functions involved with the subtest is a strength or weakness. It is the examiner’s responsibility to become actively engaged with the pattern of subtests, behavioral observations, and any other relevant sources of information necessary to determine which ability or abilities are high and low for the person. For example, Coding requires rapidly processing the digit-symbols, planning, sequencing, learning the digit-symbol pairs, making the actual response, speed, and a high level of motivation. For one person, scoring low on Coding might reflect poor speed; for another, it might reflect

difficulties with short-term memory; and for a third, it might reflect poor motivation. If a clinician noted a relatively low score on Coding, he or she might make sure that other sequencing-oriented tasks (i.e., Digit Span, Arithmetic, Letter-Number Sequencing) were also low. If this were the case, it would support the interpretation that sequencing was a problem. If this were not the case, an alternative explanation needs to be determined. Thus, a clinician might suspect that slow speed was the problem. If the examinee has done poorly due to slow speed, it would be expected that scores on other speeded tests would also be low (check Block Design, Symbol Search, and Cancellation). Sometimes behavioral observations might be useful. For example, a person scoring low on Coding due to problems with visual acuity might be struggling with seeing the images or have corrective lenses that are only partially effective.

Another example might be a clinician who is trying to decide whether the examinee prefers a simultaneous or sequential style of processing information. A relevant behavior for careful observation would be the way the examinee approached the task on Block Design. Did he or she proceed in a step-by-step sequence, trying to match each block with a segment of the picture; or, rather, did he or she try to understand the design as a whole while attempting to complete the task? A final relevant example might be low scores on Arithmetic, Digit Span, Coding, Letter-Number Sequencing, and Symbol Search. Each of these subtests requires a high level of motivation. Indeed, sometimes they have been referred to as validity indicators because scores on them are likely to be lowered as a result of poor motivation. Rather than work to decipher the examinee's low abilities as reflected in these subtests, the clinician might decide that behavioral observations more accurately suggest the person was not expending a sufficient amount of effort.

This procedure should follow statistical principles for calculating subtest strengths and weaknesses, but at the same time it should not be a rigid, mechanical process. For example, the scores on WAIS-IV/WISC-V subtests of a client who presents with subjective complaints related to poor sequencing (e.g., difficulty following directions, placing things in the wrong order) may not necessarily all be quite within the statistically interpretable weakness range (indicated as subtest "strengths" and "weaknesses" on the Record Form). However, given the quite clear symptom reports (and possibly behavioral observations), practitioners may still choose to interpret the sequencing-related subtests. In contrast, another client might have most sequencing subtests in the statistically significant range, but poor sequencing was not a symptom complaint and no behavioral observations were noted that would have been consistent with poor sequencing. As a result, the hypothesis of poor sequencing might be rejected as not applying to (or meaningful for) the person. The outlined procedure, then, should be used for hypothesis generation in which other factors beyond the mechanical interpretation procedure can confirm or disconfirm these hypotheses.

Level IV. Qualitative/Process Analysis

In addition to interpreting subtest variability, a qualitative/process approach tries to understand the underlying reasons why a score is high/low. A general strategy is to look at the content of responses, especially on Information, Vocabulary, Comprehension, and Similarities. Frequently the presence of unique, highly personal, or unusual

responses can suggest some important dimensions of an individual's intellectual or personality functioning (see Groth-Marnat et al., 2000; E. Kaplan, Fein, Morris, & Delis, 1991; E. Kaplan, Fein, Morris, Kramer, & Delis, 1999). For example, some responses may reflect aggressive tendencies, concrete thinking, or unusual associations. A highly aggressive person might provide unusual responses on some of the Vocabulary items, and a person with paranoid personality characteristics might provide rigid, cautious, and legalistic responses. Similarly, impulsivity might be suggested by persons who quickly place incorrect blocks together on Block Design and then do not reflect on whether their designs were correct.

A more formal approach is to work with the process scores. For example, providing a timed versus an untimed (No Time Bonus) administration and scoring for Block Design enables clinicians to understand the relative extent that fast or slow responses determined a client's score. These procedures are not used to help compute the Full Scale IQ, index, or additional cluster scores. In addition, they are optional procedures; they are quite time consuming and should not be routinely calculated unless there is some reason to suspect they might yield useful information. Another reason for not routinely calculating these procedures is that more research needs to be performed to better understand the meaning of norms, cutoff scores, and the degree to which they increase interpretive accuracy. However, they can be used to help generate hypotheses related to the client's functioning. The various process scores can be divided into five clusters based on which subtests they refer to.

1. Process score related to Block Design

Block Design No Time Bonus (BDN). The usual Block Design subtest score does not provide information on the extent to which speed versus difficulty with visuoconstructive abilities have impacted the score. Some individuals may have intact visuoconstructive abilities but, due to a cautious problem-solving style, slow cognitive processing, or physical difficulties, may have low scores on the usual Block Design index score. The BDN scoring reduces but does not completely eliminate the importance of speed. The reason that it does not completely eliminate the importance of speed is that, in order to obtain points, clients must still perform the task correctly within the maximum time limit. A client who does much better on BDN compared with the normal scoring is likely to have good visuoconstructive abilities but, for some reason, does not work quickly (check also the Processing Speed Index and subscale scores on Coding, Symbol Search, and Cancellation).

2. Process scores related to Digit Span

Digit Span Forward (DSF), *Digit Span Backward (DSB)*, and *Digit Span Sequencing (DSS)*. It is sometimes useful to compare a client's performance for Digit Span Forward versus Digits Span Backward. Digit Span Forward is a fairly simple, straightforward procedure. As a result, it is fairly stable in that it resists cognitive deterioration. In contrast, DSB and DSS require more attention, concentration, and ability to manipulate numbers. It is thus useful to compare scores on DSF with scores on DSB. If DSB is much lower than DSF, it suggests cognitive weakness related to poor attention, sequencing, chunking, and visualizing the numbers. It is also useful to compare DSF with DSS since, as with DSB, the

sequencing task is more difficult. A significantly lower score on DSS similarly reflects weakness in attention, sequencing, chunking, and visualizing numbers.

Longest Digit Span Forward (LDSF) versus Longest Digit Span Backward (LDSB). This score represents a variation on DSF versus DSB and is derived simply by noting the longest number of digits recalled forward versus the longest recalled backward. As such, it is a raw rather than a scaled score. The reason for evaluating the LDSF is that sometimes it is useful to double check DSF versus DSB. The interpretation is essentially the same as for DSF versus DSB.

Longest Digit Span Sequence (LDSS) versus Longest Letter-Number Sequence (LLNS). This score represents a variation on LDSF versus LDSB. It is similarly based on raw scores derived from noting the longest number of digits recalled forward versus the longest string of Letter-Number Sequencing completed. As Letter-Number Sequencing represents a more difficult variation on Digit Span Sequencing, the interpretation is comparable to LDSF versus LDSB.

3. Letter-Number Sequencing (LLNS)

The LLNS score is a raw score based on the total number of letters and numbers that were recalled on the last correct trial. This score may provide additional information on how a client performed on Letter-Number Sequencing, especially if performance on the previous items was quite variable.

4. Process scores related to Cancellation Random (CAR) versus Cancellation Structured (CAS; WISC-V only)

Item 1 of Cancellation has objects arranged in a random order. In contrast, Item 2 includes the added feature that the objects are lined up in even rows. This means that the task for Item 2 is somewhat easier than that for Item 1. It is expected that the greater structure of Item 2 would result in better performance. (See Tables C.15 and C.16 in the *WISC-V Administration and Scoring Manual* for conversion and comparisons.) The improved performance assumes, of course, that the child actually benefits from the greater structure. Thus, the extent to which the child benefits from the structure can be determined by comparing Cancellation Random (CAR) with Cancellation Structured (CS).

5. Naming Speed (NS; WISC-V only)

For children ages 6 through 8, two different types of items are administered in the Naming Speed task. Six-year-olds are asked to rapidly name colors and objects (Naming Speed Color-Object; NSco) and separately to rapidly name colors and objects along with their size (Naming Speed Size-Color-Object; NSsco). When NSco is greater than NSsco, it can indicate that the added demand of another attribute to identify overtaxed the child or that size (large/small) is not as overlearned and rehearsed as object names and colors. If NSsco is greater than NSco, it suggests that something likely interfered with the child's performance on the simpler NSco task, such as problems with motivation or attention.

Seven- to 8-year-olds are asked to do the NSsco task as well as naming letters and numbers as quickly as possible (Naming Speed Letter-Number; NSln). Similar interpretation follows discrepancy between these tasks as with NSco and NSsco. If NSsco is greater than NSln, it is likely that the child has not become

automatic in his or her letter and number identification yet. If NSIn is greater than NSsco, it suggests that something likely interfered with the child's performance on NSsco, such as attentional or motivational problems.

These process scores allow clinicians to elaborate on the meanings of individual subtests. An expanded series of subtest stimuli, alternative administrations, and scorings can be found in the *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009a).

Level V. Intrasubtest Variability

A further, potentially important area of analysis involves looking at the patterns of performance within the items of each subtest. Items in subtests are arranged in sequences that become progressively more difficult. Thus, a normal and expected pattern would have the examinee pass the initial items and slowly but evenly begin to fail more difficult ones. A more sporadic pattern, in which the examinee misses initial easier items but passes later more difficult ones, may suggest an attentional deficit or specific memory losses, particularly related to retrieval difficulties (E. Kaplan et al., 1991, 1999). If performance is highly inconsistent, the reason should be explored further. For example, clients might be consciously faking if they miss every other item, miss extremely easy items, and/or appear much more alert than their obtained IQ. Emotional difficulties may be interfering with performance sporadically as well. Sporadic performance might also be characteristic of patients with brain damage with diffuse cortical (Mittenberg, Hammeke, & Rao, 1989) or subcortical involvement (Godber, Anderson, & Bell, 2000). An analysis of the intrasubtest scatter can provide a type of information different from that obtained merely by looking at the quantitative scaled scores. It should be noted, however, that research on intrasubtest scatter is equivocal, given that J. J. Ryan, Paul, and Arb (1999) were unable to find high subtest scatter on the Information subtest among patients who had documented retrieval difficulties.

WECHSLER INDEXES AND SUBTESTS

To understand the Wechsler intelligence scales, it is useful to understand the various abilities that the indexes and subtests measure. This section describes the indexes and subtests and presents the different abilities involved in each of the WAIS-IV and WISC-V subtests, followed by a discussion of their relevant features, including the possible meanings associated with high or low scores. Subtest abilities and factor loadings for the WAIS-IV are based on research reviewed in the *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008b). Descriptions of the subtest abilities and data on factor loadings presented for the WISC-V subtests are derived from work on the WISC-IV by Kaufman and Lichtenberger (2006), Flanagan and Kaufman (2009), and Sattler (2008) as well as the *WISC-V Technical and Interpretive Manual* (Wechsler, 2014b).

In keeping with the overall approach of this book, any interpretations suggested in the discussion of the subtests should be considered tentative. They are merely beginning

possibilities that must be explored further and placed in a proper context. In addition, no subtest is a pure measurement of any single intellectual ability; rather, each represents a combination of skills. It is important to emphasize that a low or high score on a specific subtest can occur for a variety of reasons, which the examiner must consider in interpreting the overall profile. This section is most helpful only after practitioners are familiar with the subtest stimuli and administration procedures outlined in the WAIS-IV and WISC-V manuals.

Verbal Comprehension Index/Subtests

The Verbal Comprehension Index (WAIS-IV: Vocabulary, Similarities, and Information; WISC-V: Similarities and Vocabulary) is a relatively pure, refined measure of verbal abilities. The material presented to examinees is in the form of oral questions that they need to answer. An examinee's score on Verbal Comprehension reflects the extent to which he or she understands the meanings of words, can conceptualize verbal information, the extent of his or her factual knowledge related to verbal material, and his or her ability to adequately express the material in words. Thus, Verbal Comprehension measures an individual's proficiency in these areas:

- The ability to work with abstract semantic information
- The amount and degree of benefit a person has received from his or her educational background
- Verbal memory abilities
- Verbal fluency

Everyday examples of persons scoring high suggest that they will be able to easily understand spoken communication and easily construct sentences, be verbally fluent, likely do well in verbally oriented occupations, and have interests in educational activities. In contrast, persons scoring low may have difficulties with spoken language, struggle over fluidly coming up with the appropriate words to express themselves, and have little interest in educational or intellectual pursuits.

These considerations should be tempered by the fact that the subtests on the Verbal Comprehension Index are more influenced by cultural factors than other subtests. In contrast, many of the more nonverbally oriented subtests comprising most of the Perceptual Reasoning (Visual Spatial and Fluid Reasoning on the WISC-V) and Processing Speed indexes are considered to be somewhat more culture free. If an individual does significantly better (9 points or more for the WAIS-IV or 12 points or more for the WISC-V) on the Verbal Comprehension Index compared with the Perceptual Reasoning Index (WAIS-IV) or Visual Spatial Index (WISC-V), this difference may indicate a number of interpretative possibilities, including a relatively high level of education; a tendency toward overachieving; psychomotor slowing because of depression; difficulty working with practical tasks; deficits in nonverbal abilities; poor visual-motor integration; or a quick, impulsive work style resulting in relatively more errors on Perceptual Reasoning subtests (Kaufman & Lichtenberger, 2006; Lezak et al., 2012; Sattler, 2008). In addition, persons from professional occupations,

with high educational attainment, and with high IQs in general are likely to have quite high Verbal Comprehension Index scores.

Similarities (WAIS-IV/WISC-V Core Subtest)

- Logical abstract reasoning*
- Verbal concept formation or conceptual thinking
- Distinguishing essential from nonessential details
- Associative ability combined with language facility

The Similarities subtest requires verbal concept formation and abstract reasoning ability. These functions mediate an awareness of the belonging-togetherness of objects and events of the day-to-day world. An essential aspect of adjusting to one's environment is the use of these abilities to clarify, reduce, and classify the style and manner to which a response is made. Inductive reasoning is required, as the examinee must move from particular facts to a general rule or principle. Implicit in the test is the ability of individuals to use long-term memory and to apply elegant expressions in their responses. The more precise and abstract the expression, the higher the score, which indicates that verbal fluency is an important determinant. Correct responses to the last few items indicate a particularly high level of abstraction. Individuals with a good ability for insight and introspection tend to perform highly on this subtest; thus, it may be used as an indicator of favorable prognosis for psychotherapy. Scores decrease significantly in individuals with schizophrenia, rigid or inflexible thinkers, and patients with dementing conditions. Examiners can, therefore, use this subtest to gain further information regarding the nature of an examinee's idiosyncratic or pathological form of concept formation.

High scorers show good verbal concept formation, which, if unusually high, may reflect intellectualizing tendencies. Low scorers show poor abstraction abilities, literalness, and inflexible thinking. The Similarities subtest in adult protocols is the most sensitive subtest to left-hemisphere lesions, particularly lesions to the left temporal and/or left frontal regions (Dobbins & Russell, 1990).

Vocabulary (WAIS-IV/WISC-V Core Subtest)

The Vocabulary subtest includes these abilities or traits:

- Language development*
- Word knowledge*
- Expressive language ability*
- General verbal intelligence
- Language usage and accumulated verbal learning ability
- Rough measure of the subject's optimal intellectual efficiency
- Educational background
- Range of ideas, experiences, or interests that a subject has acquired

*Abilities followed by an asterisk indicate specific abilities and traits strongly associated with the subtest under discussion here through page 195.

The Vocabulary subtest is a test of accumulated verbal learning and represents an individual's ability to express a wide range of ideas with ease and flexibility. It may also involve the person's richness of ideas, long-term memory, concept formation, and language development. Vocabulary is noteworthy in that it is one of the most reliable Verbal Comprehension subtests (WAIS-IV test-retest reliability = .89; WISC-V test-retest reliability = .90) and, like Information, it is highly resistant to neurological deficit and psychological disturbance (Lezak et al., 2012; Reitan & Wolfson, 1993). Although the Vocabulary subtest holds up with age, it tends to fall off with those people for whom visuospatial skills are far more important than verbal abilities. Vocabulary generally reflects the nature and level of sophistication of the person's schooling and cultural learning. Vocabulary is primarily dependent on the wealth of the early educational environment, but it is susceptible to improvement by later experience or schooling. Vocabulary is the best single indicator of general intelligence (correlation with the Full Scale IQ of .72 on the WAIS-IV; .66 on the WISC-V). Because of its high degree of stability, Vocabulary is often used as an indicator of a person's intellectual potential and to make an estimate of the premorbid level of functioning (see subsection on estimating premorbid IQ in the "Assessing Brain Damage" section).

A qualitative analysis of Vocabulary responses, similar to Comprehension and Similarities, often provides useful information relating to the examinee's thought processes, background, life experiences, and response to frustration. It is often important to explore incorrect responses to determine whether they were guesses, clang associations (e.g., "ponder" meaning "to pound" or "assemble" meaning "to resemble"), concrete thinking, bizarre associations, or overinclusive reasoning. Even when a response is correct, a consideration of the style used to approach the word and specific content can be helpful.

High scores suggest high general intelligence and indicate that the examinee can adequately recall past ideas and form concepts relating to these ideas. Persons with high scores have a wide range of interests and a good fund of general information, and they may have high needs for achievement. Clinical populations who score high on Vocabulary may use compulsive or intellectualizing defense mechanisms. Low scores suggest a limited educational background, low general intelligence, poor language development, lack of familiarity with English, and/or poor motivation.

Information (WAIS-IV Core Subtest, WISC-V Supplemental Subtest)

- Range of general factual knowledge*
- Old learning or schooling
- Intellectual curiosity or urge to collect knowledge
- Alertness to day-to-day world
- Long-term memory

The Information subtest samples the type of knowledge that average persons with average opportunities should be able to acquire. This knowledge is usually based on habitual, overlearned material, particularly in the case of older children and adults. Both Information and Vocabulary are highly resistant to neurological deficit and

psychological disturbance (Lezak et al., 2012; Reitan & Wolfson, 1993) and are two of the most stable subtests. Because of this stability, Wechsler referred to them as “hold” tests as opposed to “no-hold” tests, which he theorized are more sensitive to deterioration and such situational variables as anxiety and fatigue (e.g., Arithmetic, Coding, Block Design). Furthermore, both these subtests are good measures of general intelligence and are highly correlated with educational level (Kaufman & Lichtenberger, 2006) and WAIS-IV and WISC-V Full Scale IQs. Research has shown that the earlier WAIS-R Information and Vocabulary subtests have predicted college grade-point average as accurately as well-established college aptitude tests (Feingold, 1983). It is for these reasons that Information (along with Vocabulary and Arithmetic) is included in Bannatyne’s Acquired Knowledge category. It also loads most strongly (.84) on the Verbal Comprehension factor.

Although performance on the Information subtest involves remote memory and alertness to the environment, it is influenced only to a small extent by conscious effort and is believed to be only minimally affected by factors such as anxiety. To score well, the individual must have been exposed to a highly varied past environment, have an intact long-term memory, and possess a wide range of interests.

A high score on this subtest suggests that the examinee has good long-term memory, cultural interests, strong educational background, positive attitude toward school, good verbal ability, and possibly intellectualization as his or her most frequently used defense mechanism. Low scorers may show superficiality of interests, lack of intellectual curiosity, cultural deprivation, lack of familiarity with Western (primarily American) culture (however, note the availability of numerous foreign-country adaptations), or poor educational opportunity. Failing initial easy items combined with success on more difficult ones (high intrasubtest variability; see Level IV procedure) may suggest difficulties with retrieval, although research substantiating this hypothesis has been equivocal (E. Kaplan et al., 1991; Mittenberg et al., 1989; J. J. Ryan & Paul, 1999). High intrasubtest scatter may also suggest the possibility of malingering or poor motivation.

Comprehension (WAIS-IV and WISC-V Supplemental Subtest)

- Demonstration of practical knowledge*
- Knowledge of conventional standards of behavior and social norms*
- Ability to evaluate past experience; that is, proper selection, organization, and emphasis of facts and relationships*
- Abstract thinking and generalization (later items only)*
- Social maturity, social judgment, common sense, or judgment in practical social situations
- Grasp of social milieu; for example, information and knowledge of moral codes, social rules, and regulations
- Reality awareness, understanding, and alertness to the day-to-day world

Comprehension has often been considered to reflect the extent to which an examinee adheres to conventional standards, has benefited from past cultural opportunities, and has a well-developed conscience. However, formal studies have generally not supported

a relationship between Comprehension and various measures of social intelligence (see Beebe, Pfiffner, & McBurnett, 2000). Comprehension is also, at least in part, a test of information, which is supported by its high correlation (low- to mid-70s, depending on age) with the Information and Vocabulary subtests. Comprehension involves an adaptive response by the individual to a situation that requires him or her to select the most efficient way of dealing with a specific problem. The examinee not only must possess relevant information but also must appropriately use this information for decision making. In this sense, the Comprehension subtest goes one step beyond the degree of complexity and synthesis required for the Information subtest. Like Vocabulary and Information, it measures general verbal ability (.79 correlation with WAIS-IV VCI, .65 correlation with WISC-V VCI). Not only must the examinee have the necessary information, but he or she also must apply it in a coherent, problem-oriented manner. Thus, a Comprehension score significantly below the Information score suggests that an examinee is not effectively using or applying his or her knowledge.

In assessing an examinee's responses, it can be important to distinguish between actually dealing with the material to develop an original response and merely repeating overlearned concepts. For example, parroting answers to "forest," "parole system," or the proverbs does not indicate full comprehension and may simply be based on past experience rather than on accurate problem solving, good judgment, or abstract reasoning. Thus, basic rule-of-thumb answers can significantly increase the total number of correct responses. However, on the later items, a correct response requires higher-level problem solving, and these items, therefore, can be good measures of general intelligence instead of merely rote memorization.

Personality variables, especially those relating to judgment, are important areas to consider on this subtest. In particular, poor levels of adjustment can lower scores on Comprehension. Clinicians should note the pattern of responses, clichés, literalness, and any circumscribed responses. In contrast, good judgment involves the ability to engage in discriminative activity. Failure on the easy items even though later, more difficult items are passed indicates impaired judgment. It is important to note emotional implications on this subtest because emotional responsiveness influences the way a person evaluates environmental events. For example, individuals who are highly analytical and use these analytical abilities to avoid emotions may have difficulty understanding the social components of situations as presented in Comprehension.

High scorers show reality awareness, capacity for social compliance, good judgment, and emotionally relevant use of information. Low scorers, especially if their scores are 4 or more subscale points below their Vocabulary scores, might have poor judgment, impulsiveness, and hostility against their environment. Persons with mental disturbances often do poorly on Comprehension, which may be the result of disturbed perceptions, idiosyncratic thinking, impulsiveness, or antisocial tendencies.

Perceptual Reasoning (WAIS-IV) and Visual Spatial (WISC-V) Index/Subtests

The Perceptual Reasoning Index on the WAIS-IV (Block Design, Matrix Reasoning, Visual Puzzles) and Visual Spatial Index on the WISC-V (Block Design, Visual Puzzles) are relatively pure measures of examinees' perceptual abilities. An examinee's score reflects the extent to which he or she has good nonverbal reasoning, can integrate nonverbal material, pays close attention to detail, and accurately responds to the

visuospatial material presented to him or her as well as his or her nonverbal fluid reasoning on the WAIS-IV. Many of these abilities involve using the kind of visuospatial and visuomotor skills to solve problems that are not taught in formal academic schooling. The Perceptual Reasoning Index and Visual Spatial Index reflect proficiency in these areas:

- The individual's degree and quality of nonverbal contact with the environment
- The ability to integrate perceptual stimuli with relevant motor responses
- The capacity to work in concrete situations
- The ability to evaluate visuospatial information

Everyday examples of persons who score high in Perceptual Reasoning/Visual Spatial include good ability to follow maps, accurately driving from one place to the next, correctly assembling objects, finding objects in a house/office, drawing a design, and the ability to work with nonverbal material. In contrast, low scores suggest that the person may have a difficult time following spatial directions, finding objects placed in a house/office, accurately drawing designs, repairing broken objects, or estimating distance.

Perceptual Reasoning/Visual Spatial subtests are generally less affected by educational background than are the Verbal Comprehension subtests. If an individual does significantly (.05 level) better (9 points or more on the WAIS-IV, 12 or more points on the WISC-V) on the Perceptual Reasoning Index or Visual Spatial Index compared with the Verbal Comprehension Index, this may indicate a number of interpretive possibilities, including superior perceptual organizational abilities, a tendency toward low academic achievement, possible acting out (juvenile delinquency), an individual who could be described as a doer rather than a thinker, a person from a relatively low socioeconomic background, presence of a language deficit, poorly developed auditory conceptual/processing skills, or that (specifically with the WAIS-IV PRI) immediate problem solving is better developed than problem solving based on accumulated knowledge.

Block Design (WAIS-IV/WISC-V Core Subtest)

- Analysis of whole into component parts*
- Spatial visualization*
- Nonverbal concept formation
- Visuomotor coordination and perceptual organization
- Capacity for sustained effort; concentration
- Visual-motor-spatial coordination; manipulative and perceptual speed

The Block Design subtest involves nonverbal problem-solving skills because it emphasizes analyzing a problem into its component parts and then reintegrating these parts into a cohesive whole. The examinee must apply logic and reasoning in a manner that will solve spatial relationship problems. As a test of nonverbal concept formation, Block Design demands skills in perceptual organization, spatial visualization, and abstract conceptualization. The Block Design subtest is sturdy and reliable,

correlating highly with general intelligence, and is not likely to be lowered except by the effects of depression or organic impairment. It also has been found to relate to everyday measures of spatial abilities (Groth-Marnat & Teal, 2000). To perform well, examinees must be able to demonstrate a degree of abstraction that is free from literal concreteness. They must also make a distinction between part and whole by demonstrating both analytic and synthetic skills. This test involves an ability to shift the frame of reference while maintaining a high degree of flexibility. The examinee must also be able to inhibit impulsive tendencies and to persist in a designated task.

An important feature of Block Design is that it enables an examiner to observe the examinee's responses. Some subjects are easily discouraged and give up, while others insist on completing the task even if they have to work beyond the time limit. In approaching the task, one subject might impulsively place the blocks together in a non-random sequence, whereas another subject might demonstrate a meticulous sequential style, thereby revealing preferences for either a holistic simultaneous or a more sequential problem-solving style. Additional observations can reveal factors such as hand preference, motor coordination, speed of information processing, frustration tolerance, and ability to benefit from feedback. A highly reflective or compulsive style can lower scores because of the resulting extended time for completing the task. Placing blocks outside the 2×2 or 3×3 configuration is a further behavioral observation that reflects poor visuospatial skills (Kramer, Kaplan, & Huckleba, 1999). Thus, potentially valuable information can be obtained by observing and recording differences in solving the Block Design tasks.

Block Design is also a nonverbal, relatively culture-free test of intelligence. It is reliable in that it correlates highly with general intelligence (.66 correlation with the WAIS-IV Full Scale IQ, .73 correlation with the WISC-V Full Scale IQ), but it has a relatively low correlation with education. Thus, the Block Design subtest is only minimally biased by an examinee's cultural or educational background. Block Design scores can therefore be important tools in assessing the intellectual potential of persons from diverse cultural and intellectual backgrounds.

Block Design is an excellent indicator of right-hemisphere brain damage and is especially sensitive to right parietal lesions (Lezak et al., 2012; Reitan & Wolfson, 1992, 1993). Patients with right-hemisphere lesions tend to make errors because they might distort the designs, misperceive aspects of them, or become disoriented when attempting to complete them. In contrast, patients with left-hemisphere lesions, particularly if the lesion is in the parietal lobe, are not nearly as likely to have a poor Block Design score. However, when they do, it is likely to be expressed in design simplification, confusion, and a concrete approach to reproducing the design (Lezak et al., 2012). Inattention (neglect) can be reflected by the examinee's failing to complete the right or left portion of the design. For example, only 6 or 7 of the blocks might be used when attempting to complete a 9-block design (Lezak et al., 2012). Scores on Block Design are typically among the lowest subtests in patients with Alzheimer's disease. It is sensitive to the early phases of the disease and thus can be useful in differentiating between Alzheimer's and pseudodementing conditions such as depression (Fuld, 1984; La Rue & Jarvik, 1987).

High scorers show a good capacity for visuospatial perception, visuomotor speed, a good ability to concentrate, and excellent nonverbal concept formation. Low scores

suggest poor perceptual abilities, difficulties with visual integration, and problems in maintaining a sustained effort. It is important to note that, as a timed subtest, psychomotor slowing (or other problems with processing speed) can lower scores on Block Design, which would not reflect problems with perceptual or visual integration abilities.

Visual Puzzles (WAIS-IV Core Subtest, WISC-V Core Subtest for Index Calculation, but Not FSIQ)

- Visual recognition and identification*
- Perception of the parts in relation to the whole*
- Visuospatial reasoning*
- Analysis of wholes into component parts
- Capacity for sustained visual effort; concentration

Visual Puzzles was a new subtest developed for the WAIS-IV. It requires examinees to first view a completed design at the top of a page. They must then look at six possible design fragments at the bottom of the page and select the three that, when combined, would re-create the original design at the top of the page. It is thus similar to solving a puzzle. However, examinees do not actually manipulate any objects. The task is performed completely based on visual reasoning. Examinees must analyze the original shape and then synthesize which of the alternative shapes are correct. It is thus a measure of visual and analogical reasoning (Wechsler, 2008b). The *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008b) reported that similar measures have been found to involve “visual perception, broad visual intelligence, fluid intelligence, simultaneous processing, spatial visualization and manipulation, and the ability to anticipate relationships among parts” (p. 14). However, some research has indicated that it is not a pure measurement of visual perceptual ability, as mental flexibility, processing speed, and even language abilities are related to performance (Fallows & Hilsabeck, 2012).

High scores indicate good nonverbal reasoning, ability to concentrate, and a good ability for visuospatial integration. In contrast, low scores suggest problems with integrating nonverbal material, possible problems with concentration, and possible visual neglect.

Picture Completion (WAIS-IV Supplementary Subtest, Not Available on WISC-V)

- Visual alertness*
- Visual recognition and identification (long-term visual memory)*
- Awareness of environmental detail; reality contact
- Perception of the whole in relation to its parts; visual conceptual ability
- Ability to differentiate essential details from nonessential details
- Visual concentration combined with an ability to visually organize material

The Picture Completion subtest is a measure of visual concentration and is a non-verbal test of general information. It involves discovering consistency and inconsistency by paying close attention to the environment and accessing remote memory. It is dependent on, and also draws on, an individual's experience with his or her culture.

Thus, persons who are unfamiliar with common features of American/Western society may make errors because of a lack of experience rather than a lack of intelligence. Persons will also make errors if they are unable to detach themselves emotionally from the material, which will thereby make accurate discriminations difficult. For example, passive, dependent personalities might make errors because they notice the absence of people controlling the actions in the pictures. Typical responses might be that “there’s nobody holding the pitcher,” “there are no people rowing the boat,” or “there’s no flagpole.” Sometimes negative, inflexible, oppositional individuals state that there is nothing missing in the pictures.

High scorers are able to recognize essential visual information, are alert, and demonstrate good visual acuity. Low scores indicate poor concentration and attention to detail and inadequate visual organization. Impulsiveness can often produce lowered performance because the examinee may make a quick response without carefully analyzing the whole picture.

Perceptual Reasoning (WAIS-IV) and Fluid Reasoning (WISC-V) Index/Subtests

The Fluid Reasoning Index (WISC-V: Matrix Reasoning, Figure Weights) is based on research showing that some subtests are equally (or near equally) influenced by both perceptual reasoning and working memory (L. C. Ward, Bergman, & Hebert, 2012). The WAIS-IV includes the content of the WISC-V Fluid Reasoning Index in its Perceptual Reasoning Index, though only the Matrix Reasoning subtest is included in the PRI calculation. An examinee’s score on the Fluid Reasoning Index reflects the extent to which he or she has good nonverbal, fluid reasoning; can identify underlying patterns among nonverbal stimuli; can use the relationships among different stimuli to determine a single, underlying concept; and applies themes, common concepts, and underlying factors to provide solutions to problems. The Fluid Reasoning Index reflects proficiency in these areas:

- The individual’s ability to solve problems not based on previous knowledge
- Abstract reasoning
- Inferential reasoning, classification of stimuli into conceptual categories, and concept formation
- The ability to apply implicit patterns and constructs to visuospatial information

Everyday examples of persons who score high in Fluid Reasoning include solving novel problems that do not rely on processes usually employed, good facility in mathematics, creative and novel thinking, and being adaptable in situations that present roadblocks. In contrast, low scores reflect comfort with more concrete and expected tasks, using familiar processes; clear guidance and direction; and straightforward thinking.

Like Visual Spatial subtests, Fluid Reasoning subtests are generally less affected by educational background than are the Verbal Comprehension subtests. If an individual does significantly (.05 level) better (12 or more points on the WISC-V) on the Fluid Reasoning Index compared with the Verbal Comprehension Index, this may indicate a number of interpretive possibilities, including superior novel problem-solving abilities, a tendency toward low academic achievement, an individual who could be described as

a doer rather than a thinker, a person from a relatively low socioeconomic background, presence of a language deficit, poorly developed auditory conceptual/processing skills, or that immediate problem solving is better developed than problem solving based on accumulated knowledge.

Matrix Reasoning (WAIS-IV/WISC-V Core Subtest)

- Visuospatial reasoning*
- Abstract reasoning*
- Visual organization*
- Simultaneous processing of visuospatial information
- Analysis of wholes into component parts

High scores on Matrix Reasoning suggest good visual information processing and nonverbal abstract reasoning skills. On the WAIS-IV, Matrix Reasoning is combined with Block Design and Visual Puzzles to form the Perceptual Reasoning Index. In contrast, the WISC-V uses Matrix Reasoning on the Fluid Reasoning Index. Matrix Reasoning is untimed (though the instructions say to score an item incorrect if an individual does not respond within approximately 30 seconds) and is therefore useful for persons from older age groups who might do poorly on some of the other timed tests. It also does not penalize those who have a reflective, cautious problem-solving style. Matrix Reasoning is relatively culture-free and requires only a minimal amount of visuomotor coordination because the subject merely points to the correct response. Conceptually, Matrix Reasoning is similar to the Halstead Reitan Category Test and Raven's Progressive Matrices. However, studies have not determined the nature and degree of correspondence between these measures.

One of the rationales for developing the Matrix Reasoning subtest was to develop a visuospatial subtest with good psychometric properties that could replace the psychometrically poor Object Assembly subtest. Matrix Reasoning has been found to have test-retest stabilities ranging from .74 to .92, SEM of .95 to .99, and a factor loading of .73 on the Perceptual Reasoning Index on the WAIS-IV and a factor loading of .67 on the Fluid Reasoning Index on the WISC-V. It is an excellent measure of general intelligence (.67 correlation with WAIS-IV Full Scale IQ and .72 correlation with the WISC-V Full Scale IQ).

High scores might indicate good nonverbal abstract reasoning abilities, a preference for simultaneous processing of information, and excellent visual information processing. Low scores might suggest low visual concept formation, poor or at least rigid visual reasoning, or poor concentration. Negativism might be indicated if the examinee seems unmotivated and replies with wording such as "none of them match."

Figure Weights (WAIS-IV Supplementary Subtest, WISC-V Core Subtest)

- Nonverbal mathematical reasoning*
- Quantitative and analogical reasoning*
- Visual concentration combined with an ability to visually organize material
- Capacity for sustained effort

Figure Weights is a relatively new subtest developed for the WAIS-IV and WISC-V. Examinees are first shown a picture at the top of the stimulus card depicting two measuring scales with various shapes on the trays. The scales on the left are balanced, as indicated by the two trays being at the same level. The scales on the right are also balanced in that they are similarly at the same level. There are objects in the left tray but the right tray is empty, with a question mark appearing above the empty tray. The examinee is then requested to select which of five sets of objects at the bottom of the stimulus card would be necessary to balance the scale depicted at the top right side of the stimulus card (and with the question mark on top of the tray). Figure Weights is a nonverbal test of fluid and mathematical reasoning. It involves quantitative and analogical reasoning and inductive or deductive logic. To a certain extent, the task is similar to Arithmetic in that it involves sustained effort and concentration. However, the Arithmetic subtest requires examinees to hold verbal details of the problem in their memory and work with this information (verbal working memory). In contrast, Figure Weights has the problem depicted visually. As a result, the importance of remembering the various components of the problem is minimized.

High scores suggest good nonverbal quantitative reasoning, excellent concentration, and a good ability to organize nonverbal information. In contrast, low scores might indicate poor nonverbal quantitative reasoning and difficulties with concentration. A potentially useful contrast is to compare scores on Arithmetic with those on Figure Weights (see WAIS-IV, Table B.3, and WISC-V, Table B.7). If scores are the same, it suggests a person's quantitative abilities are equal for both verbal and nonverbal information. However, a significantly higher Arithmetic score (3 or more subscale points) suggests that verbal quantitative abilities are better developed. In contrast, a relatively higher Figure Weights score (3 or more subscale points) indicates that the examinee's nonverbal quantitative abilities are better.

Picture Concepts (Not Available on WAIS-IV, WISC-V Supplemental Subtest)

- Nonverbal concept formation*
- Perceptual recognition
- Abstract, categorical reasoning

In Picture Concepts, clients are initially requested to look at two rows of pictures. They are then asked to determine which picture in the first row goes together with one of the pictures from the second row (e.g., they might both be animals or means of transportation). Later items use three instead of two rows. Clients must first scan the pictures and determine which categories the pictures belong to. They must then decide which is the most important characteristic that the picture on the first row shares with the picture on the second row. Picture Concepts is thus a measure of nonverbal concept formation and abstract reasoning. It functions as a nonverbal parallel to the verbally oriented Similarities subtest.

Picture Concepts is a somewhat weak measure of overall intelligence, with a low to moderate correlation with the WISC-V Full Scale IQ ($r = .48$). It is a moderate contributor to the Fluid Reasoning Index (loading = .54) and has ample specificity. It is

moderately reliable ($r = .83$) and actually has its highest correlation with Vocabulary ($r = .42$).

In order to interpret Picture Concepts, it is often useful to note if there are any visual difficulties or signs of negativism. It is also important to note whether clients respond in a rapid, impulsive manner or one that is more reflective. Once the subtest is completed, it is sometimes useful to develop a better understanding of the reasoning behind clients' problem solving. This can be done by asking why they made the response they did. It might also be useful to check whether clients have performed much higher or lower on the Similarities subtest. Scores that are relatively higher on Picture Concepts suggest that abstract reasoning is better for nonverbal than for verbal material. In contrast, lower scores on Picture Concepts indicate that nonverbal abstract reasoning is superior.

High scores suggest good nonverbal concept formation, good flexibility of thinking, and good logical and abstract thinking. Low scores indicate poor nonverbal concept formation, rigid thought processes, and poor abstract reasoning. However, sometimes low scores might reflect good abstract reasoning but the responses are based on more unconventional, innovative means of perceiving relationships between pictures. Questioning clients about their responses might help determine whether their answers were wrong due to their not understanding the problem or simply due to them understanding it correctly but developing unconventional responses.

Arithmetic (WAIS-IV Core Subtest, WISC-V Supplemental Subtest)

Although the Arithmetic subtest is categorized within the Fluid Reasoning Index on the WISC-V, it is categorized within the Working Memory Index on the WAIS-IV. See the next section, "Working Memory Index/Subtests," for a description of this subtest.

Working Memory Index/Subtests

Working Memory (WAIS-IV: Digit Span, Arithmetic, Letter-Number Sequencing; WISC-V: Digit Span, Picture Span) is a more complex and controversial construct than the constructs measured in the other indexes. It has been related primarily to concentration, attention, and short-term memory. Because the term *memory* is in the title, clinicians sometimes are tempted to think of it as a measure of memory. However, working memory is not the same as memory. Instead, it is a narrow measure of the ability to hold and manipulate information for a short period of time. Whereas this ability is certainly related to and a prerequisite for many aspects of memory, it is not the same as memory. As such, clinicians should never interpret high (or low) scores as indicating that the client has good (or poor) "memory." In addition, on the WAIS-IV, the subtests focus on auditory/verbal aspects of working memory rather than on visual components (see Leffard et al., 2006); this is not the case for the WISC-V, which includes both verbal and nonverbal measures of working memory. Sequencing is also crucial for Working Memory subtests because each subtest requires that the respondent place numbers and symbols in their proper order. Wielkiewicz (1990) has suggested that the low scores on Working Memory can reflect not only poor concentration, memory, and sequencing but also difficulties with executive functioning. Specifically, the person experiences difficulty attending to stimuli and

simultaneously performing other mental tasks (e.g., listening to spoken digits and storing them while simultaneously reversing them and then repeating them backward). Good performance also requires a high level of motivation. As a result of these diverse functions, a poor working memory is also likely to lower performance in other areas, and this should be considered when estimating the person's overall potential. Thus, Working Memory represents a proficiency in these areas:

- Concentration and attention*
- The ability to hold and manipulate information in short-term memory*
- Short-term memory
- Sequencing
- Facility with numbers (especially on the WAIS-IV)
- Mental flexibility (especially for Digit Span Backward, Digit Span Sequencing, and Letter-Number Sequencing)

Persons scoring high would be expected to be good at recalling phone numbers, evaluating checks in a restaurant, following a sequence of instructions, concentrating on a task without being distracted, and “multitasking” (easily doing two things at the same time). In contrast, low scores might indicate that the person would have a difficult time paying attention to a lecture, recalling phone numbers, following a sequence of instructions, and doing two things at the same time, or the person might be quite anxious.

It is crucial to consider a variety of interpretive possibilities to interpret the Working Memory Index. Often behavioral observations can be crucial. A client who frequently asks to have the questions repeated might have a high level of distractibility. Alternatively, a high degree of motor activity or excessive talking might highlight a client's high level of anxiety. If number skills have not been developed, the client might ask to write out the numbers related to the arithmetic problems or count out the numbers with his or her fingers.

The association between Working Memory and ADHD has been equivocal. Conceptually, it would seem that Working Memory would be the lowest index score (Mayes & Calhoun, 2007). However, often this has not been found (Reinecke et al., 1999; Wechsler, 2008b, 2014b). As mentioned previously, this may be due partially to the structured nature of the testing situation, which eliminates most distractions. In contrast, a real-world environment typically has many competing distractions that a person with ADHD may have a difficult time screening out. As a result, clinicians need to look at additional sources of information, including corroborating sources and rating scales (e.g., Connors's Rating Scales, Child Behavior Checklist, Behavior Assessment System for Children).

Digit Span (WAIS-IV/WISC-V Core Subtest)

- Immediate rote recall*
- Reversibility; ability to shift thought patterns (from digits forward to digits backward to digit sequencing)*
- Concentration and attention

- Auditory sequencing
- Rote learning

Digit Span is considered to be a test of short-term memory and attention. The examinee must recall and repeat auditory information in the proper sequence. Bannatyne (1974) further described this as “auditory vocal sequencing memory.” Correct responses require a two- and three-step process (three steps for Digit Span Backward and Sequencing). First, the information must be accurately received, which requires attention and encoding. Persons who are easily distractible have difficulty in this phase. Second, for Digit Span Backward and Sequencing, the examinee must manipulate and resequence the information. Finally, for all three Digit Span trials, the examinee must accurately recall and vocalize the information. Persons who can perhaps receive the information correctly may still have a problem at the recall phase if they have short-term memory difficulties because they cannot hold the memory trace long enough. Sometimes the previous digit is forgotten as they are attempting to vocalize a present one. Digit Span Forward is a simpler, more straightforward task requiring rote memory than Digit Span Backward and Digit Span Sequencing. During Digit Span Backward and Digit Span Sequencing, the examinee must hold the memory longer and also transform it before making a restatement. Thus, a good performance on Digit Span Backward and Digit Span Sequencing is likely to reflect a person who is mentally flexible, can concentrate, and is tolerant of stress. High Digit Span Backward and Digit Span Sequencing scores may also involve the ability to form, maintain, and scan visual mental images formed from the auditory stimulus (Lezak et al., 2012; Sattler, 2008; Wielkiewicz, 1990). Digits Sequencing has been found to be particularly sensitive to the impact of Alzheimer’s disease and traumatic brain injury (Wechsler, 2008b). A Digits Forward score that is 4 or more points higher than the Digit Span Sequencing score is significant and should be investigated further (see WAIS-IV Administrative Scoring Tables C.2, C.3, C.4, pp. 247–249 and WISC-V Administrative Scoring Supplement, Tables C.14, C.15, C.16, and C.17, pp. 98–120).

Passive, anxiety-free individuals seem to do best on the Digit Span task. It requires an effortless and relatively unhampered contact with reality, which is characterized by open receptivity to incoming information. Performance is greatly hampered by increased anxiety or tension, and the Digit Span subtest is considered the most susceptible to the effects of anxiety. Other subtests that also are sensitive to the effects of anxiety are Arithmetic, Coding, and Letter-Number Sequencing. On the WAIS-IV, Digit Span and Arithmetic form the Working Memory Index and are generally (along with the Processing Speed subtests) the most sensitive subtests to brain damage, mental retardation, and learning disabilities (Lezak et al., 2012; Wechsler, 2014b). Similarly, the Digit Span subtest (and Picture Span) is included in the WISC-V Working Memory Index.

Persons who score high have good auditory short-term memory and excellent attention and may be relatively unaffected by stress and anxiety. However, just because a person has good short-term auditory memory for digits does not necessarily mean that his or her memory for more complicated information, such as music or verbally relevant information, is also good. These more complex features of memory may have to be assessed by other means. The rare event of Digit Span Backward being longer than

Digit Span Forward (.9% of adult WAIS-III profiles, .2% of children's protocols; Wechsler, 2003b) suggests that the individual has excellent numerical abilities. Low scores on Digit Span indicate difficulty concentrating, which may be the result of anxiety, unusual thought processes, or a host of other influences, such as lack of sleep or hunger. A large discrepancy (5 digits) in favor of Digit Span Forward versus Digit Span Backward can suggest the presence of an organic deficit, particularly if the overall backward Digit Span score is below scores for tests such as Information and Vocabulary. Whereas Digit Span Forward is fairly stable and resistant to deterioration, Digit Span Backward and Digit Span Sequencing are far more difficult than Digit Span Forward and are quite sensitive to deterioration (see subsection on estimating premorbid IQ in the "Assessing Brain Damage" section). Whereas Digit Span Forward is more likely to be lowered by left-hemisphere lesions, lowered Digit Span Backward is more consistent with either diffuse or right-frontal involvement (Lezak et al., 2012; Rapport, Webster, & Dutra, 1994). Lowered performance for both Digit Span Backward and Coding occurs with the diffuse damage associated with exposure to solvents (Groth-Marnat, 1993; Morrow, Furman, Ryan, & Hodgson, 1988).

Picture Span (Not Available on WAIS-IV, WISC-V Core Subtest for Index Calculation, but Not FSIQ)

- Nonverbal working memory*
- Perceptual patterning*
- Attention, concentration
- Short-term visual memory
- Rote learning with proactive interference

On this subtest, clients are shown a page with a number of pictures on it for a short period of time. They are then shown another page with more pictures, some of which were present on the first page and some of which were not. They are asked to identify which pictures were on the first page, in order from left to right. They receive points for correctly identifying which pictures were present and extra points for identifying them in the correct order. Later items include more pictures and more distractors. Notably, pictures are reused on different items on the task, which can cause proactive interference. That is, information from an earlier task can "interfere" with the learning on a new task; children can learn what pictures are present on a page, but when asked to identify them on the next page, they may accidentally "remember" pictures that were not on that first page, because they were on a previous item. This kind of proactive interference has been shown to increase working memory demands on an individual (Blalock & McCabe, 2011; Carroll et al., 2010). Although on the surface this subtest seems to function as a nonverbal analogy to Digit Span (especially Digit Span Forward), it is much more likely that the learning uses both visual and verbal rehearsal, linking the verbal and nonverbal working memory processes.

Picture Span is a moderate estimate of overall intelligence, with a low to moderate correlation with the WISC-V Full Scale IQ ($r = .53$). It is a large contributor to the Working Memory Index (loading = .64) and has ample specificity. It is moderately reliable ($r = .85$) and has its highest correlation with Digit Span ($r = .51$).

In order to interpret Picture Span, it is often first useful to note if there are any visual difficulties or signs of poor motivation. The task takes a great deal of targeted, focused, purposeful attention and effort, so poor motivation can easily impact the score negatively. Poor motivation can present as “I don’t know” responses as well as careless, hasty responding. It can be useful to compare the results of Picture Span to Picture Concepts, to see if visual spatial perceptual issues are present, as well as to Digit Span, to see if there are marked differences in verbal and nonverbal working memory.

Although Picture Span is a newer subtest and not much research has been conducted on it, high scores suggest good attention and concentration, nonverbal short-term memory, and patterning and sequencing abilities. Low scores indicate possible difficulty in one or more of these areas. Because performance on this task requires a sequence of motivation, attention and concentration, short-term visual memory, and sequencing and inhibition (because of the proactive interference), a low score on Picture Span will not easily specify where in this sequence the client had difficulty. Behavioral observations can help the assessor determine whether motivation was a factor. Additionally, it can be helpful to determine qualitatively whether the client remembered the pictures but not in order, lowering his or her score, or if he or she performed poorly because of not remembering the pictures at all. This can help distinguish problems with short-term visual memory from problems with nonverbal patterning and sequencing. Other subtests and measures can help clinicians further determine where problems may have occurred in the series of abilities necessary to successfully complete the task.

Arithmetic (WAIS-IV Core Subtest, WISC-V Supplemental Subtest)

- Computational skill*
- Auditory short-term memory
- Sequencing ability
- Numerical reasoning and speed of numerical manipulation
- Concentration and attention/low distractibility
- Reality contact and mental alertness (i.e., active relationship to the outside world)
- School learning (earlier items)/acquired knowledge
- Logical reasoning, abstraction, and analysis of numerical problems (later items)

The Arithmetic subtest requires focused concentration as well as basic mathematical skills and an ability to apply these skills. The basic skills required to complete this test are usually acquired by the time a person reaches junior high school; therefore, low scores are more likely to be the result of poor concentration. Arithmetic is likely to be more challenging and stressful than subtests like Information and Vocabulary, because the task itself is more demanding, because the test is timed, and because of widely felt negative associations with mathematics. Thus, persons who are susceptible to the disruptive effects of anxiety are likely to be adversely affected. However, examiners may want to establish whether the person simply lacks the necessary skills or had difficulty concentrating. This can be assessed by readministering items that the examinee had previously missed, but allowing the person to use paper and pencil without a time limit.

Under these circumstances, persons with adequate mathematical knowledge who are distractible should be able to complete the items correctly.

Individuals from higher socioeconomic backgrounds, obedient teacher-oriented students, and persons with intellectualizing tendencies usually do well on this subtest. A helpful formula is that Information plus Arithmetic equals school achievement. Because numbers come from the outside environment and create rule and direction, some individuals react rebelliously. This is particularly true for antisocial personalities. Histrionic personalities, who do not readily accept outside direction and generally refuse to take responsibility for their behaviors, may likewise do poorly. This is not to suggest that lowered Arithmetic scores are diagnostic of these clinical groups but rather that this lowering may at times be consistent with the way these individuals interact with their environment.

High scorers show alertness, capacity for concentration, freedom from distractibility, and good short-term auditory memory, and they may use intellectualizing defenses. Low scorers show poor mathematical reasoning, lack of capacity to concentrate, distractibility, and poor auditory short-term memory. A poor educational background in which adequate mathematical skills have not been developed can also account for lowered performance.

Letter-Number Sequencing (WAIS-IV/WISC-V Supplemental Subtest)

- Auditory short-term memory*
- Sequencing ability*
- Concentration and attention*

A good performance on Letter-Number Sequencing suggests that the person has good sequencing, attention, and concentration. The subtest requires the person to attend to a series of letters and numbers that have been read to him or her, hold them in short-term memory, manipulate them into a new order, and repeat the new sequence. Psychometrically, Letter-Number Sequencing is good to adequate. WAIS-IV test-retest reliability has been found to be .80, the SEM is 1.03, and it has a factor loading of .69 with the Working Memory Index. On the WISC-V, it has been found to have a test-retest reliability of .82 (for all ages), the SEM is 1.13, and it has a factor loading of .79 with the Working Memory Index.

High scores suggest that the examinee has good short-term auditory memory, is able to sequence auditory information effectively, is persistent, and has good working memory. In contrast, a low score indicates the person has difficulties with auditory sequencing, has poor short-term auditory memory, is inattentive, and may also be impulsive, anxious, or poorly motivated.

Processing Speed Index/Subtests

The Processing Speed Index (PSI; WAIS-IV/WISC-V: Symbol Search, Coding) reflects the mental and motor speed with which a person can solve nonverbal problems. Further subtest support for this index can be found if the person also has correspondingly high (or low) performances on the timed nonverbal Block Design subtest. In addition to mental and motor speed, the Processing Speed factor is a measure of a person's

ability to plan, organize, and develop relevant strategies. Because speed and concentration require good test-taking attitudes, Processing Speed (as well as Working Memory) can be lowered by poor motivation to perform well. For this reason, these two indexes are sometimes referred to as *validity* factors. Whether a lowered performance is the result of poor motivation is often best assessed by behavioral observations in combination with clarification and consideration of the presenting problem. An overly reflective problem-solving style could also lower the Processing Speed factor because the person would take too much time cautiously considering his or her response to each item. Thus, Processing Speed represents proficiency in these areas:

- Speed of processing information
- Planning and organization
- Motor control and coordination of visual and motor abilities
- Motivation

People who score high can generally solve problems quickly, are likely to be fast readers, can dial a telephone number rapidly, can quickly find a telephone number in a phone book, may do well in occupations that require rapid responding, and would be able to quickly locate food items on a shelf. In contrast, persons who score low may require extra time to learn material, be slow picking up objects, be slow readers, be hesitant, carefully reflect on their answers before giving them, or take a relatively long time to find food items on a shelf.

Processing Speed is noteworthy in that it is the index that is most sensitive to cognitive problems caused by a wide variety of disorders, including dementia, traumatic brain injury, ADHD, and learning disabilities (Wechsler, 2008b). It can also be significantly affected by depression, for which psychomotor slowing is a specific symptom. It is also the index that begins to decrease the earliest (during a person's 20s), with a more precipitous drop in the mid-30s (Kaufman & Lichtenberger, 2006). The other indexes, particularly Verbal Comprehension, are more resistant to the effects of aging. Low scores on Processing Speed can also reflect poor motor control and be associated with problems with sensory acuity. Among persons with high IQs and those labeled as gifted, Processing Speed is often the lowest index (with Verbal Comprehension being the highest; Wechsler, 2008a).

Coding (WAIS-IV/WISC-V Core Subtest)

- Psychomotor speed*
- Ability to follow directions*
- Clerical speed and accuracy*
- Visual short-term memory*
- Paper-pencil skills*
- Ability to learn an unfamiliar task; capacity for learning and responding to new visual material
- Some degree of flexibility; ability to shift mental set
- Capacity for sustained effort, attention, concentration, and mental efficiency

- Associative learning and ability to imitate newly learned visual material
- Sequencing ability

Visuomotor integration is implied by good performance on Coding. However, the most important functions necessary for a high score are psychomotor speed combined with good recall for the symbol-digit pairs. This test involves appropriately combining the newly learned memory of the digit with the symbol, as well as adequate spatial-motor orientation, followed by executing the half-habituated activity of drawing the symbol. The subtest also requires the ability to learn an unfamiliar task, accuracy of eye-hand coordination, attentional skills, short-term memory, and the ability to work under pressure. This is a delicate and complex interaction, which can be disturbed because of difficulties with any of the preceding skills. In contrast to Vocabulary, which is a highly stable subtest, Coding is extremely sensitive to the effects of either organic or functional impairment. In particular, patients with depression and those with brain damage have a difficult time with this subtest. It is also the subtest that is most influenced by age. For example, a WAIS-IV raw score required to achieve a subscale score of 10 for the 70- to 74-year-old group would obtain a subscale score of only 6 when compared with the 20- to 34-year-old reference group.

Coding pairs with Symbol Search to form the Processing Speed Index. Coding is a fair measure of general intelligence (.59 correlation with the WAIS-IV Full Scale IQ, .50 correlation with the Full Scale IQ on the WISC-V).

Because visuomotor coordination (particularly visual acuity and motor activity) is implied, it is not surprising to find that those individuals with high reading and writing experience are among the high scorers. Functions that are implicit in the task are rapid visual, spatial, and motor coordination, as well as the executive action of drawing the symbol. Because this task requires sustained attention and quick decision making, anxious hesitancy, obsessiveness, deliberation, and perfectionism significantly lower scores. This difficulty might be somewhat counteracted by informing persons who appear perfectionistic and reflective that they need only make their response legibly, but it does not need to be perfect. Persons who become highly anxious in competitive situations may also be adversely affected. Coding scores can be lowered by anxiety; the psychomotor slowing found in depressive states or the confused orientation of schizophrenia likewise produces a decrease in performance. Thus, a rough index of the severity of a person's depression can be assessed by comparing the relative lowering of Coding with other more stable subtests. Of particular significance is that Coding is one of the most sensitive subtests to the effects of any type of organic impairment (Lezak et al., 2012; Reitan & Wolfson, 1993; Wechsler, 2008b), and it tends to be one of the lower scores found in individuals with learning disabilities (Bannatyne, 1974; Groth-Marnat, 2001; Kaufman & Lichtenberger, 2006). Even in persons with minimal brain damage, Coding is still likely to be the lowest subtest overall (Lezak et al., 2012; Reitan & Wolfson, 1993). In addition, patients with rapidly growing tumors are more likely to have lower scores than those with slow-growing tumors (Reitan & Wolfson, 1993).

Because Coding requires such a diverse range of abilities, high or low scores can potentially indicate a wide number of possibilities. Therefore, clinicians need to work particularly hard to extract the significance of scores by integrating them with other

relevant measures, behavioral observations, and medical/personal history. A measure like the Bender–2 may help determine if motor, perceptual, or visual-motor integration problems may be the cause of low Coding scores.

High scorers potentially have excellent visuomotor ability, mental efficiency, capacity for rote learning of new material, and quick psychomotor reactions. Lower scorers may have reduced capacity for visual associative learning, impaired visuomotor functioning, and poor mental alertness.

Symbol Search (WAIS-IV Core Subtest, WISC-V Core Subtest for Index Calculation, but Not FSIQ)

- Speed of visual search*
- Speed of processing information
- Planning
- Encoding information in preparation for further processing
- Visuomotor coordination
- Learning ability

Symbol Search was designed to be as pure a test as possible of information-processing speed, without the motor demands present in Coding. It pairs nicely with Coding because, conceptually, they assess similar areas, as is more formally indicated by relatively high correlations (WAIS-IV, .65; WISC-V, .58) between the two subtests. Together, they form the Processing Speed factor. Symbol Search is psychometrically a relatively good subtest. Test-retest over an 8- to 82-day ($M = 22$) interval was .81 for the WAIS-IV; over a 9- to 82-day ($M = 26$) interval, it was .80 for the WISC-V. It correlated relatively highly with the Full Scale IQ on the WAIS-IV (.64) but not as highly on the WISC-V (.46), and it correlated highly with the Processing Speed (WAIS-IV, .91; WISC-V, .89) composites.

High scores suggest that the individual can absorb information rapidly as well as integrate and respond to this information. In addition, it suggests good levels of visuomotor coordination, short-term visual memory, planning, general learning, and a high level of attention and concentration. Low scores suggest slow mental processes; visual-perceptual difficulties; possibly poor motivation and/or anxiety; difficulties with short-term visual memory; or a reflective, perfectionistic, or obsessive problem-solving style.

Cancellation (WAIS-IV/WISC-V Supplemental Subtest)

- Perceptual recognition*
- Perceptual discrimination*
- Perceptual scanning ability*
- Speed and accuracy*
- Attention and concentration*
- Visuomotor coordination

On this subtest, the examinee is shown a page with a few selected shapes or pictures in one area. A larger number of shapes/pictures include both the original

shapes/pictures as well as a large variety of other shapes/pictures. Examinees are requested to scan the area with the wide variety of shapes/pictures and then draw a line through only those shapes/pictures from the selected few shapes/pictures. Thus, examinees must scan the shapes/pictures, recognize the correct ones, screen out the distracting ones, make the motor task of crossing them out, remember what they are supposed to be looking for, and maintain a constant level of attention. It is important that examinees make the discriminations as quickly as possible. The WAIS-IV has increased the difficulty of this task by requiring examinees to identify both shape and color. The subtest consists of two items that are in the form of two sets of pictures. One item consists of a haphazard arrangement of pictures (“random arrangement”), and the second item consists of pictures in neat rows (“structured arrangement”). Most of the time it is sufficient to calculate the combined scores on these two items. However, sometimes it can be useful to obtain separate scores for the random versus the structured arrangement (see Cancellation Random/CAR versus Cancellation Structured/CS process scorings in the *WISC-V Administration and Scoring Manual*).

The *WAIS-IV Technical and Interpretive Manual* reported that the Cancellation subtest is similar to other subtests that measure “processing speed, visual selective attention, vigilance, perceptual speed, and visual motor ability” (Wechsler, 2008b). Within neuropsychological contexts, low scores on cancellation-type subtests have been found among patients with visual neglect, motor perseveration, and difficulties inhibiting their responses (Lezak et al., 2012).

The main advantage of the Cancellation subtest is that it has a fairly high loading on the Processing Speed Index on the WAIS-IV (.56), although it has only moderate loading on the WISC-V (.41). As would be expected, the correlation between Cancellation and the Processing Speed Index is moderate (.49 for the WAIS-IV, .36 for the WISC-V). In contrast, it has quite low correlations with the other indexes and with the Full Scale IQ (.26 for the WAIS-IV, .22 for the WISC-V). Thus, it is a poor measure of *g*. It is fairly reliable (.79 test-retest on the WAIS-IV, .82 test-retest on the WISC-V) and has high specificity for all ages.

Cancellation is a particularly useful subtest for identifying the distractibility that typically occurs with ADHD or traumatic brain injury. It is often quite useful to determine whether low scores on Cancellation are due to distractibility, slow speed, or disturbances with visual recognition.

High scores indicate excellent processing speed, good attention and concentration, good perceptual recognition, good scanning abilities, good inhibition of impulses, and high motivation. Conversely, low scores suggest slow processing speed, distractibility, poor perceptual recognition, poor scanning abilities, impulsivity, and poor motivation.

Complementary Indexes/Subtests

The Complementary Indexes, Naming Speed and Symbol Translation, were added to evaluate processes that are important in the assessment of learning difficulties, including reading, general learning, and mathematics disabilities.

Naming Speed (WISC-V Only)

- Verbal and semantic speed and fluency*
- Recognition speed*
- Concentration and attention
- Automatic naming skill

Naming Speed is a new subtest on the WISC-V and is considered to be a test of automaticity of object, color, size, and letter/number recognition and naming (depending on the age). The examinee must identify and name colors and objects (age 6), size and objects (ages 6–8), or numbers and letters (ages 7–16) as quickly and accurately as possible, resulting in both a Literacy (accuracy) score and a Quantity (speed) score. The former is sensitive to general learning difficulties, and the latter is sensitive to math disabilities (Pauly et al., 2011; Willburger, Fussenegger, Moll, Wood, & Landerl, 2008).

Symbol Translation (WISC-V Only)

- Verbal-visual association memory*
- Novel learning, including short-term memory, long-term memory, and recognition*
- Concentration and attention
- Connection between verbal and visual abilities
- Motivation

Symbol Translation is the other new test on the WISC-V and is considered to be a test of learning and memory of verbal-visual associations. The examinee is taught a series of verbal-visual pairs, with one symbol representing one word, and asked to learn and remember the translation. Three separate trials assess immediate memory (short-term learning ability), delayed memory (after 20–30 minutes), and delayed recognition. The purpose of separating the learning/memory task into these three stages is to assess three different processes related to learning: learning information, which is the process assessed in the Immediate phase, focused on how well the examinee understands and learns the material in the moment; encoding information, which is the process assessed in the Recognition phase, focused on cued recognition of what was previously learned; and retrieval/recall of information, which is the process assessed in the Delayed phase, focused on the ultimate skill of having learned, encoded, and then freely recalling the information. Taken together, if there is a problem with memory (specifically learning and memory related to visual-verbal association), this index can help practitioners pinpoint where in the memory process there are problems. For example, if a client scores average on the Immediate and Recognition subtests but scores poorly on the Delayed subtest, the problem is not with learning or encoding the information but actually with the retrieval of that information from memory. Alternatively, if the client scores adequately on Immediate

but poorly on both Delayed and Delayed Recall, it is likely that he or she did not adequately encode the information and store it in memory. It is important to note that motivation is a key factor in a novel and complicated task like Symbol Translation.

ASSESSING BRAIN DAMAGE

General Principles

The Wechsler intelligence scales measure many abilities that are likely to be lowered by brain damage. These include memory, learning, perceptual organization, fluid reasoning, problem solving, and abstract reasoning. As a result, the Wechsler intelligence scales are typically a core feature of any neuropsychological battery (Groth-Marnat, 2000b; Lezak et al., 2012). At one time, it was hoped that the Wechsler intelligence scales, along with other more specialized psychological tests, could be used in the actual diagnosis of brain damage. Despite some noteworthy success in this area, it is currently more typical for psychological tests to be used in the assessment of the effects a known lesion is likely to have on a person's cognitive and adaptive functioning. Thus, the Wechsler intelligence scales, along with other specific tests of neurocognitive ability, are not tests specifically sensitive to specific brain damage. Rather, they are tests that can reflect the effects of brain damage, as well as a variety of other conditions.

During the earlier development of the WAIS and WISC, Wechsler (1958) hoped that brain damage could be discriminated based on relative lowerings in subtests that were most sensitive to neurological impairment. He referred to these brain-sensitive tests as *no-hold* tests (Digit Span, Digit Symbol/Coding, Similarities, Block Design) and contrasted them with *hold* tests, which were believed to be far more resistant to impairment (Information, Object Assembly, Picture Completion, Vocabulary). Although the distinction between hold and no-hold tests has some truth, the use of such a distinction in diagnosing brain damage has been found to result in too many misclassifications. Vogt and Heaton (1977) summarized the reasons for this lack of success by pointing out:

- There is no single pattern of brain damage, so highly variable test responses can be expected.
- The hold/no-hold distinction does not account for other significant factors, such as the age when the brain damage occurred, environmental variables, education, location of the lesion, and whether the lesion is recent versus chronic.
- The Wechsler intelligence scales do not measure many important abilities related to brain damage.

More recent work supports the theory that there is no specific brain damage profile (Groth-Marnat et al., 2000; Lezak et al., 2012). Some persons with brain damage produce low IQ scores, whereas for others, IQs are still high. Sometimes there is a high level of subtest scatter; at other times, scores on the subtests and indexes are quite even. Thus, brain damage may cause a general lowering on all or most subtests; at other times, there may be a lowering of only specific subtests. The most general indicator for the detection of brain damage is whether a person's scores (either general or

specific) are lower than expected given his or her socioeconomic status, age, education, occupation, and other relevant areas of his or her history.

One of the older conventional wisdoms about brain damage is that left-hemisphere involvement is more likely to lower the verbal abilities, whereas right-hemisphere involvement results in relatively lower scores on nonverbal abilities. This finding would be reflected on the WAIS-IV in that patients with left-hemisphere damage would be expected to have relatively lower scores on the Verbal Comprehension Index than scores on the Perceptual Reasoning Index ($VCI < PRI$). In contrast, patients with right-hemisphere damage would be expected to have lower scores on Perceptual Reasoning than Verbal Comprehension ($PRI < VCI$; see previous discussion under “Level II: Indexes and Additional Groupings”). Research using earlier versions of the Wechsler intelligence scales evaluating Verbal versus Performance IQs found that sometimes this laterality effect occurred; at other times, it did not (Aram & Ekelman, 1986; R. A. Bornstein, 1983; Kaufman & Lichtenberger, 2002, 2006; Larrabee, 1986; Lezak et al., 2012). In general, right-hemisphere lesions are likely to produce greater verbal-nonverbal discrepancies than left-hemisphere lesions (see review by Kaufman & Lichtenberger, 2006). Probably the safest approach is that a VCI – PRI difference is not diagnostic either of brain damage in general or, more specifically, of damage to one hemisphere or the other. However, a significant VCI – PRI difference can at times be consistent with this hypothesis. Specifically, a lowered Verbal Comprehension score suggests the possibility of language impairment, whereas a relatively lower Perceptual Reasoning Index suggests right-hemisphere impairment. Clinicians should avoid overinterpreting a person’s results and should consider other means of investigation, including knowledge of health status, medical history, and additional specialized psychological tests.

A finding noted in the *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008b) is that brain damage is most likely to impact a person’s processing speed (see also Dikmen et al., 2009; Fisher, Ledbetter, Cohen, Marmor, & Tulskey, 2000; K. A. Hawkins, 1998). Patients with both traumatic brain injury and Alzheimer’s disease all had Processing Speed as their lowest score. In contrast, verbal abilities were better preserved. This finding reflects the fact that persons with brain impairment tire easily and have difficulties with concentration and attention. In addition, Perceptual Reasoning and Working Memory were lower among patients with brain injuries. Perceptual Reasoning, like Working Memory and Processing Speed, involves fluid intelligence. From a theoretical perspective, fluid intelligence is tied more to an intact brain structure and also is assessed more clearly by the ongoing problem-solving tasks presented in the Perceptual Reasoning subtests. Thus, a destruction of brain tissue would be more likely to lower fluid intelligence, which would be reflected in lowered Perceptual Reasoning subtest scores. This hypothesis can be further assessed by calculating the CHC subtest groupings for fluid intelligence (see “WAIS-IV/WISC-V Successive-Level Interpretation Procedure” section, Level IIb).

Many of the inferences related to brain damage depend on profile analysis. Useful material relevant to brain damage can be found in the discussion of Levels II through V under the “WAIS-IV/WISC-V Successive-Level Interpretation Procedure” section in this chapter and in the relevant discussions for each subtest in the “Wechsler Indexes and Subtests” section. Much of this interpretation depends on hypothesis testing in

which the practitioner integrates knowledge about the person, brain function, Wechsler subtests, and other information. Often no clear, empirically based guidelines exist. Accuracy of any inferences is based partially on whether they make neuropsychological sense. However, one generally accepted principle is that intersubtest scatter is most likely to occur with focal lesions of recent origin (see Kaufman & Lichtenberger, 2006). In contrast, a general lowering of all abilities (low subtest scatter) is more likely with either chronic lesions or diffuse degenerating diseases (e.g., exposure to neurotoxins; Groth-Marnat, 1993; L. Miller, 1993).

One useful strategy developed by E. Kaplan and her colleagues is to work toward parceling out the underlying processes responsible for scores on the Wechsler intelligence scales (Milberg et al., 1996). Alternative administration guidelines, error categories, useful tables, and interpretive procedures have been developed for the WAIS-IV/WMS-IV (*Advanced Clinical Solutions*, Pearson, 2009a). For example, a clinician might be interested to know if a client's poor performance on Information or Vocabulary resulted from lack of knowledge or problems with retrieval. This might be determined by presenting him or her with multiple-choice formats that assist with the retrieval process (recognition of correct answers rather than the much more difficult task of having to recall them). If a client does significantly better on the multiple-choice format than the standard format, it suggests that the lowering was caused by retrieval difficulties. The WAIS-IV has incorporated some of these strategies into the process scores. For example, the Digit Span subtest has separate scores for Digit Span Forward, Digit Span Backward, and Digit Span Sequencing. Digit Span Backward and Digit Span Sequencing are more difficult than simply repeating the digits forward. As a result, the backward and especially the sequencing tasks are more sensitive to patients with brain injuries in that they obtain fairly low scores on these versions when compared to the overall score on the Digit Span subtest (Wechsler, 2008b). Another strategy built in to the process approach is to carefully investigate various error categories (Groth-Marnat et al., 2000; E. Kaplan et al., 1991, 1999). For example, visual neglect might be indicated by not noticing details on the left (usually) side of pictures on Picture Completion or making errors on the left side of the designs for Block Design.

When the preceding strategies, principles, and cautions are taken into account, clinicians can generate and test useful hypotheses developed from different patterns of subtest scores. The next list summarizes some of the most frequently supported hypotheses about specific subtests or patterns of subtests:

- The Processing Speed Index, along with its related subtests (Symbol Search and Coding), is the most brain-sensitive Wechsler index and can be lowered by lesions in any location. A lowering implies difficulties with speed of information processing and/or learning, sequencing, rote learning, concentration (especially with lowerings on Digit Span and Arithmetic), and visuomotor abilities (K. A. Hawkins, 1998; Lezak et al., 2012; Reitan & Wolfson, 1992; Wechsler, 2008b, 2014b).
- Block Design is also brain sensitive, especially to either left or right parietal lesions (Golden, 1976; Lezak et al., 2012; McFie, 1960, 1969). A lowering implies visuospatial problems and possible difficulty in constructing objects

(constructional apraxia: note quality of drawings; Kramer et al., 1999; Zilmer, Bell, Fowler, Newman, & Stutt, 1991). The Block Design No Time Bonus process score removes the timed variable. If the score is still low, hypotheses that visuospatial problems are present are strengthened (Wechsler, 2008b).

- Both Digit Span and Arithmetic are frequently lowered in populations with brain damage, particularly with left-hemisphere lesions (Kaufman & Lichtenberger, 2006; Lezak et al., 2012; McFie, 1960, 1969). Lowering suggests poor concentration and attention and, if Digits Backward is significantly lower than Digits Forward (generally 5 or more digits), a significantly reduced level of mental flexibility and/or difficulty forming and maintaining a visual image of the digits. It may also suggest difficulties in a person's executive functions related to selecting a key stimulus, attending to it, and maintaining the information in short-term storage while simultaneously performing other mental tasks (Wielkiewicz, 1990).
- Vocabulary, Information, and Picture Completion have often been used as a rough estimate of a person's premorbid level of functioning because they are relatively unaffected by lesions. An important exception is that children with brain damage often score lowest on the Vocabulary subtest (Boll, 1974; Reitan, 1974; Reitan & Wolfson, 1992). In addition, Information and Vocabulary are generally lowered (especially relative to Similarities) in patients with left temporal damage, suggesting difficulties with word comprehension, retrieval, and language expression (Dobbins & Russell, 1990). Picture Completion, while usually resistant to brain damage, might be lowered because of difficulties involving vision, especially visual agnosia (difficulty recognizing objects; E. Kaplan et al., 1991, 1999). Thus, always considering Vocabulary, Information, and Picture Completion as indicators of premorbid functioning can potentially result in incorrect inferences. As a result, scores on these three subtests should be interpreted in relation to what is known about brain-behavior relationships.
- The Similarities subtest, especially in relation to Information and Vocabulary, is most likely to be lowered with left-frontal lesions and suggests difficulty with verbal reasoning and verbal concept formation (Dobbins & Russell, 1990).
- Qualitative responses, particularly related to error categories (even when the subtests are not lowered), can provide useful information related to brain damage. Some responses might suggest poor judgment and impulsivity, whereas others might indicate concrete thinking in which the person is bound by the stimulus value of the item (e.g., a season of the year might be described as "hot, dry" rather than the more abstract reference to a season; or the clang response to a word such as "empathy" being defined as "empty"). Other persons might report they once knew the answer but have forgotten, which can be assessed through multiple-choice options. Diffuse brain damage (but not focal) might also be consistent with a high degree of intratest scatter in which the client misses easy items but correctly answers later, more difficult ones (Mittenberg et al., 1989). This finding suggests retrieval failure and/or the random loss of previously stored information. This intrasubtest scatter is most likely to occur on Vocabulary, Comprehension, Information, Similarities, and Picture Completion.

Estimating Premorbid IQ

Professional psychologists are frequently confronted with the need to estimate a client's premorbid level of functioning (functioning before an injury or illness that has potentially caused cognitive deterioration of some sort). In an ideal situation, previous IQ test results derived before the injury could be obtained and compared with the current level of functioning. Even in this situation, clinicians should be aware that a decline in *overall* performance should not be inferred unless there is a significantly lower current IQ than had been obtained from a premorbid IQ assessment. A discrepancy of 12 or more Full Scale IQ points on the WAIS-R when compared with a Full Scale IQ on the WAIS-III was found to result in an 80% accurate detection of adults who had actually suffered a cognitive decline (Graves, Carswell, & Snow, 1999). It should also be stressed that there still might be quite specific areas of decline that are not sensitive to the global measure of IQ scores.

In most cases, premorbid IQ results are not available; therefore, clinicians must rely on other strategies to infer premorbid ability. These strategies include historical achievement-based records, current measures of ability that are not sensitive to decline ("hold" measures), demographic-based regression equations, or a combination of these. Useful historical records might include grade-point average, SAT or ACT scores, work achievement records, achievement tests, or peer ratings (see Baade & Schoenberg, 2004). The age of the person, as well as relevant aspects of the injury (e.g., size and location of the lesion, recency of injury), might also be important to consider.

A further strategy for estimating premorbid ability is to note performances on Wechsler subtests that are considered most resistant to neurological impairment (Information, Picture Completion, and especially Vocabulary). As discussed previously, these subtests have often been considered to reflect the person's past level of functioning and have therefore been referred to as hold subtests. Administering an achievement test such as the Wide Range Achievement Test (WRAT4) or Wechsler Individual Achievement Test (WIAT-III) might accomplish a similar purpose. One difficulty is that for many clients, especially those who are well educated, this method is likely to overestimate premorbid IQ. In contrast, this method would likely underestimate premorbid IQ for subgroups whose premorbid Performance Scales are typically greater than Verbal Scales (e.g., Native Americans, Hispanics, bilingual individuals, persons with low educational attainment, blue-collar workers). It would also underestimate premorbid ability among patients who have had impairment to their verbal abilities.

A related technique is to consider the person's two or three highest subtests on the WAIS-IV (regardless of whether the subtests are brain sensitive or non-brain sensitive) and then use these to estimate the person's premorbid level of functioning. Despite its occasional usefulness, this procedure is likely to result in a high number of misclassifications because it does not consider crucial factors such as the person's age, educational level, or location of the lesion (Matarazzo & Prifitera, 1989). In addition, it is quite common for normal persons, especially those with high IQs, to have significant variability in their performance. As a result, estimates of the premorbid level of functioning of clients with high IQs are likely to be significantly inflated.

A variation of this hold procedure is to use a reading test, such as the National Adult Reading Test (NART; H. Nelson & Williams, 1991) or Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). The NART and WTAR were designed by selecting 50 irregularly spelled words (e.g., yacht, naive) that are unlikely to be pronounced correctly unless the client has previous knowledge of the words. This relatively pure recognition task places minimal demands on memory and problem-solving abilities. A NART-estimated WAIS-R Full Scale IQ 20 points higher than a person's obtained IQ was shown to suggest intellectual decline (80% accuracy for those with actual decline; Graves et al., 1999). The WTAR has been found to correlate highly with other measures of premorbid functioning and to be quite stable despite cognitive changes measured by other tests (K. Green et al., 2008). However, this stability assumes that the injury would not have affected the person's reading ability. The NART and WTAR both have calculations that take into account some demographic variables, though they obviously cannot take into account all possible variables that might relate to premorbid functioning.

Other efforts to determine premorbid IQ have used regression equations based on demographic variables (education, occupation, etc.). One of the most extensively researched is the Barona Index (Barona, Reynolds, & Chastain, 1984). To correctly classify (80% accuracy) clients with true cognitive decline, a discrepancy of 25 IQ points would be required (Graves et al., 1999). Unfortunately, this discrepancy is sufficiently large such that other more straightforward procedures (e.g., previous work performance, grade-point average, medical records) would be likely to be more accurate. In addition, the index is likely to be inaccurate for persons with either extremely high (above 120) or extremely low (below 69) IQs (Barona et al., 1984; Graves et al., 1999; Veiel & Koopman, 2001), and the formulas are likely to overestimate most premorbid IQ levels (Eppinger, Craig, Adams, & Parsons, 1987).

Research with the WAIS-IV found a series of algorithms that were useful in estimating premorbid IQ. An example that can be used for general clinical purposes is listed next (Oklahoma Premorbid Intelligence Estimate, WAIS-IV [OPIE-4]; Holdnack, Schoenberg, Lange, & Iverson, 2013):

OPIE-4 (2 subtests) using Vocabulary and Matrix Reasoning

$$\begin{aligned} \text{FSIQ} = & 57.847 + .707 (\text{Vocabulary}) + 1.387 (\text{Matrix Reasoning}) \\ & - .204 (\text{Age in years}) + .0000415 (\text{Age in years cubed}) \\ & - 4.231 [\text{if Education Kindergarten} - 7\text{th grade}] \\ & - 1.881 [\text{if Education } 8 - 10\text{th grade}] \\ & + 1.765 [\text{if Education Bachelors degree or higher}] \\ & + 1.569 (\text{Sex}) + 1.061 [\text{if lives in Midwest}] \\ & - 5.143 [\text{if African American}] \end{aligned}$$

Sex: 0 = female, 1 = male

Note: Only raw scores should be used for the subtest scores.

Holdnack et al. (2013) clarified that alternate equations should be used if the Vocabulary and Matrix Reasoning scaled scores are discrepant (see Chapter 5 Appendix in Holdnack et al., 2013). They also include equations for estimating premorbid IQ using other measures, such as the Wechsler Abbreviated Scale of Intelligence–2 (WASI-2).

Research with children has found that a series of algorithms based on the WISC-IV standardization sample has been able to predict premorbid IQ (Schoenberg, Lange, Brickell, & Saklofske, 2007), though no such equation has yet been developed for the WISC-V. This algorithm was found to be the most accurate and used a combination of demographics and raw scores on Vocabulary, Matrix Reasoning, Information, and Picture Completion:

Estimating Premorbid IQ from WISC-IV Subtests and Demographics

$$\begin{aligned} \text{IQ} = & 81.64 + 0.682 (\text{Vocabulary}) + 1.088 (\text{Matrix Reasoning}) \\ & + 0.719 (\text{Information}) + 0.317 (\text{Picture Completion}) - 4.729 (\text{Age in years}) \\ & + 0.413 (\text{Parental education}) + \text{Ethnicity} + \text{Gender} \end{aligned}$$

Parental education: Number of years of parental education (if two parents, use the average of both parents)

Ethnicity: Caucasian (0), African American (–1.646), Hispanic (0.126), Asian (3.475)

Gender: Male (0), Female (3.127)

Note: Only raw scores should be used for the subtest scores.

The total mean IQ using this formula was 99.73 with a standard deviation of 13.53, and the correlation between the estimated Full Scale IQ and the actual WISC-IV Full Scale IQ was .86 (Schoenberg et al., 2007). Fifty percent of estimated IQ scores were ± 5 points from the actual Full Scale IQs and 82% were within ± 10 IQ points. Predictions were less accurate for children in the Superior range of intelligence (>129) or at the Borderline or below range (<80). These data may not apply directly to the WISC-V, but it is hoped that work is being done to develop similar algorithms for the latest version of the test. One caution is that the preceding data were developed from healthy persons without any known neuropsychological impairment. Whether these predictions will be as accurate for clinical populations has not yet been determined. However, research using similar procedures on patients with known neuropsychological impairment for the WAIS-R, WAIS-III, and WISC-III found that the algorithms made accurate estimates (Axelrod, Vanderploeg, & Schinka, 1999; Schoenberg et al., 2003; Vanderploeg, Schinka, Baum, Tremont, & Mittenberg, 1998). Thus, it is likely that future research with clinical populations using the WISC-V and similar algorithms will provide accurate predictions.

Due to the considerable legal implications combined with the error rate, estimating premorbid IQ has been a controversial procedure (see Veiel & Koopman, 2001). These review points seem crucial:

- The equations should be used to supplement but not replace a careful evaluation of crucial information, such as work history and medical records.

- Formal cutoffs should be used. Rarely, for example, would an obtained IQ 5 to 10 points below the estimated “premorbid IQ” suggest actual cognitive decline in a person’s *overall* ability. However, this still does not preclude the possible presence of quite specific deficits (e.g., in facial recognition, short-term visual memory).
- The likelihood of errors increases when equations based on demographics or subtests are used with persons with IQs suspected of being extremely high or extremely low (<80 or >120).

Alzheimer’s Disease

The initial symptoms of Alzheimer’s disease are apathy, a decline in short-term memory, and difficulties with problem solving. Underlying these changes are reductions in cholinergic activity. Currently, neuropsychological assessment, particularly with the Wechsler intelligence scales, is one of a variety of diagnostic procedures to increase the accuracy of diagnosis as well as to understand the nature and extent of a patient’s difficulties. Cognitive assessment is often used in conjunction with medically based diagnostic procedures, such as identifying the presence of an Alzheimer’s dementia autosomal mutation among direct relatives, biomarkers in the cerebrospinal fluid, medial temporal lobe atrophy based on magnetic resonance imaging (MRI), and patterns of brain activity noted on positron emission tomography (PET) functional neuroimaging (Dubois et al., 2007).

Ideally, a unique cognitive pattern would identify the presence of Alzheimer’s disease. While a clear, distinct pattern has not been identified, research has found that nonverbal abilities seem to be more sensitive to impairment than verbal abilities. Earlier research with the WAIS-R found that a full 52% of patients with Alzheimer’s disease had Verbal greater than Performance IQ scores of 15 points or more (Fuld, 1984). Similarly, the WAIS-IV Verbal Comprehension Index ($M = 86.2$) has been found to be 10 points higher than Processing Speed ($M = 76.6$) among a sample of patients with probable mild Alzheimer’s (Wechsler, 2008b). Thus, Processing Speed is the index that is most sensitive to the early presence of Alzheimer’s disease. However, Processing Speed is also the index that is most sensitive to many if not most types of cognitive impairment, as well as emotional interference. Additional lowerings in index scores were found in Working Memory ($M = 84.3$) and Perceptual Reasoning ($M = 85.8$). The subtests that were most sensitive to being lowered by mild Alzheimer’s disease were Symbol Search, Information, Coding, and Arithmetic.

The WAIS-IV and its predecessors are useful in the assessment of Alzheimer’s disease in that they assess a wide range of general as well as specific cognitive functions. Many of these functions will be lowered as a result of Alzheimer’s disease and other forms of dementia. However, crucial functions are not measured by the Wechsler intelligence scales. Memory is one of the most important features of Alzheimer’s disease, yet the Wechsler intelligence scales do not measure this function in much depth or breadth. As a result, specialized memory scales, such as the WMS-IV, are crucial. Other functions that might need to be assessed through additional instruments include word naming (e.g., Boston Naming Test; Goodglass & Kaplan, 1983), verbal fluency (e.g., Controlled Oral Word Association Test; Benton & Hamsher, 1989), and executive

functions (e.g., Delis-Kaplan Executive Function System/DKEFS; Delis, Kaplan, & Kramer, 2001). In addition, specialized dementia batteries, such as the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery (Morris et al., 1989), have been assembled based on tests that have been found to be the most sensitive to Alzheimer's disease.

ASSESSING ADDITIONAL SPECIAL POPULATIONS

Learning Disabilities

Learning disabilities make up a complex, heterogeneous, loosely defined disorder with a wide variety of manifestations and many different theories regarding causation (Kaufman & Lichtenberger, 2006; Sattler, 2008). A central component of all definitions is that learning disabilities involve difficulties in developing skills in reading (most commonly), writing, listening, speaking, reasoning, spelling, or math. The individual diagnosed with a learning disability should have academic achievement (in at least one of the above-listed areas) that falls below what would be expected, given general ability and educational opportunity. Along with specific academic deficit, there is generally some sort of problem with information processing, such that there is likely some cognitive weakness within the intellectual profile. Further essential features are these: persons with learning disabilities have adequate intelligence, show a significant discrepancy between achievement and intellectual ability, and have a disorder that is considered primarily intrinsic to the person, presumably because of central nervous system (CNS) dysfunction. The underachievement cannot be primarily the result of an intellectual disability (formerly called mental retardation), brain damage, behavior problems, sensory disabilities, or environmental disadvantage.

The major purpose of learning disability assessment is to identify a client's strengths and weaknesses to enable the clinician to decide on an appropriate placement and to design an optimal program. Relevant areas to assess include developmental-cognitive processes, achievement, environmental demands, reactions of others to the client's difficulties, and the possible interaction of additional factors, such as fear of failure, overall level of interpersonal adjustment, and family history of similar difficulties. The Wechsler scales are typically considered essential means of identifying the client's overall level of functioning and specific cognitive strengths and weaknesses and to eliminate the possibility of intellectual disability. Other tests are usually required: achievement tests, measures of adaptive behavior, visuomotor tests, assessments of auditory and visual processing, and measures of emotional and behavioral problems, for example (see Sattler, 2008). Most often, achievement tests like the WIAT-III or K-TEA-3 (both easily interpretable with the WISC-V) are given, and two major analyses can happen: ability-achievement discrepancy analysis (comparing WISC-V scores to achievement test scores) and patterns of strengths and weaknesses (completely within the WISC-V). A discussion of some common patterns of strengths and weaknesses on Wechsler intelligence scales is presented next. For a more detailed discussion of these conceptual issues, see: Fiorello, Hale, and Snyder (2006); Flanagan, Alfonso, Mascolo, and Hale (2010); and J. B. Hale, Wycoff, and Fiorello (2011).

Researchers have expended considerable effort in searching for a specific Wechsler scale profile that is unique to populations with learning disabilities. Research on the WAIS-III revealed evidence for an ACID profile (comprised of Arithmetic, Coding/Digit Symbol, Information, and Digit Span); 24% of those diagnosed with learning disabilities had a partial (three out of the four subtests as the lowest scores) ACID profile, and 6.5% had a full (all four of the subtests as the lowest) ACID profile (Wechsler, 1997a). This is higher than the standardization sample. The WAIS-III index scores of Working Memory and Processing Speed (compared to Perceptual Organization and Verbal Comprehension) were also found to be particularly low among a sample of adults diagnosed with reading disabilities (Wechsler, 1997a). This finding has led Kaufman and Lichtenberger (1999, 2006) to suggest the possible utility of combining the five subtests in these lowest indexes into a SCALD profile (Symbol Search, Coding, Arithmetic, Letter-Number Sequencing, Digit Span).

The WAIS-IV was considerably revised compared to the WAIS-III, and these profiles are not likely to be as relevant. For example, Letter-Number Sequencing on the WAIS-IV is categorized as a supplementary subtest. However, the *WAIS-IV Technical and Interpretive Manual* reported that the lowest index score for persons with a Reading Disorder was for Working Memory ($M = 88.9$). Subtest scores were lowest for Letter-Number Sequencing, Arithmetic, and Vocabulary. These findings are consistent with the finding that persons with learning disabilities have problems with sequencing and attention.

The ACID profile also received some support with the WISC-III; most studies found that approximately 20% of persons with learning disabilities had either a partial or full ACID profile (Kaufman, 1994; Kaufman & Lichtenberger, 2002; Mayes, Calhoun, & Crowell, 1998; Stanton & Reynolds, 1998). A somewhat similar WISC profile substitutes the Symbol Search subtest for Information, resulting in the SCAD (Symbol Search, Coding, Arithmetic, Digit Span) profile. These four subtests emphasize the functions of speed of information processing, visual short-term memory, and visuomotor coordination (Symbol Search and Coding) as well as number ability and sequencing (Arithmetic and Digit Span). These functions are specifically the types that many individuals with learning disabilities (as well as many other types of brain dysfunctions) have difficulty with. Accordingly, children with learning disabilities and ADHD have been found to score particularly low on the SCAD profile (Kaufman, 1994; Mayes et al., 1998; Stanton & Reynolds, 1998). Similarly, children diagnosed with ADHD performed relatively poorly on the WISC-III Freedom from Distractibility factor (Anastopoulos, Spisto, & Maher, 1994), which included subtests similar to the Working Memory Index; on the WISC-V, children with ADHD performed worst on Coding, Arithmetic, and Picture Span as well as Picture Concepts (Wechsler, 2014b). This finding should be used with caution, however, because a relatively large proportion of children with ADHD still do not have this profile. In addition, S. B. Ward, Ward, Hatt, Young, and Mollner (1995) did not find support for the SCAD profile among children with learning disabilities.

Research with the WISC-V has not yet evaluated the ACID and SCALD profiles. However, the lowest index score among persons with learning disabilities on the WISC-IV were reported to be for Working Memory ($M = 87.0$; Mayes & Calhoun, 2007; Wechsler, 2003b). The lowest subtest scores for persons with Reading Disorders

were for Vocabulary, Letter-Number Sequencing, Information, and Arithmetic. On the WISC-V, students with Reading Disorder were similarly found to have Working Memory as their lowest index score ($M = 87.8$), with Picture Span and Vocabulary as their lowest subtest scores (Wechsler, 2014b). As expected, children with a diagnosed Mathematics Disorder had their lowest index as Fluid Reasoning ($M = 82.2$), with Arithmetic and Figure Weights as their lowest subtest scores.

Another way to understand learning disabilities and related disorders is to use Bannatyne's categories, which conceptualize learning-disabled performances as highest on subtests requiring spatial abilities (Block Design, Picture Concepts, Picture Completion) in which little or no sequencing is required (Bannatyne, 1974). Conceptual skills are intermediate (Comprehension, Similarities, Vocabulary), and subtests requiring sequencing abilities (Digit Span, Coding, Picture Span) are lowest. Thus, among many persons with learning disabilities, spatial abilities are believed to be greater than conceptual abilities, which, in turn, are greater than sequential abilities. A fourth category, Acquired Knowledge (Information, Arithmetic, Vocabulary), is also sometimes used as a rough index of the extent to which the person has accumulated school-related facts and skills. Even though these findings might suggest a greater degree of subtest scatter among persons with learning disabilities, research has not provided clear support for this finding (Greenway & Milne, 1999).

Collectively, the preceding profiles suggest that many persons with learning disabilities perform best on tasks requiring holistic, right-brain, simultaneous processing (Picture Concepts, Block Design) and worst on those requiring sequential processing (Digit Span, Coding, Picture Span), which is expressed in difficulties with planning, reading, and numerical ability. Wielkiewicz (1990) further suggested that these subtests indicate a poorly functioning executive ability in which the individual experiences difficulty attending to stimuli while simultaneously performing other mental tasks.

Reviews and cross-validation of Bannatyne's and ACID/SCAD profiles have produced inconsistent results (see Groth-Marnat, 2001). Only some groups of students with learning disabilities in some studies showed the Bannatyne Spatial > Conceptual > Sequential pattern (A. Kaufman, 1994; Kaufman & Lichtenberger, 2006). This is not surprising, given the many different modes of expression found under the umbrella term *learning disabilities* (Kaufman & Kaufman, 2001). In addition, Bannatyne's pattern has not been found to be unique to learning disabilities; it frequently occurs in a diverse number of groups, including juvenile delinquents and children with emotional disabilities (see Groth-Marnat, 2001). Although only minimal support exists for Bannatyne's categories as a diagnosis for learning disabilities, they are far from useless. The four categories (Spatial, Conceptual, Sequential, Acquired Knowledge) can be invaluable for interpreting relative strengths and weaknesses for persons with learning disabilities as well as for other groups. Although research has not been able to produce a unique profile of people with learning disabilities, the research invested in this effort has resulted in a useful means of analyzing Wechsler scale profiles.

Given the previous research, these conclusions are warranted (adapted from Groth-Marnat, 2001):

- The Full Scale IQ can be used most appropriately in the assessment of persons with learning disabilities to estimate their overall potential and assist in excluding intellectual disability as a possible explanation for poor academic performance.

- There is moderate to equivocal evidence that some profiles (relatively low Processing Speed and Working Memory, Spatial > Conceptual > Sequential, ACID, SCAD, SCALD) occur more frequently in populations with learning disabilities compared to the general population. However, these patterns need to be updated and validated for the WAIS-IV and WISC-V.
- These profiles are not unique to learning disabilities but often occur in other groups as well (e.g., juvenile delinquents, persons with ADHD or emotional disabilities).
- If a person does have a “learning-disabled” Wechsler profile (ACID, etc.), it is *consistent with, although not necessarily diagnostic of*, a learning disability.
- The majority of persons with learning disabilities *do not* have Wechsler “learning disabled” profiles. Thus, the absence of one of the profiles *does not exclude* a diagnosis of a learning disability.
- The various patterns of Wechsler subtests can, at times, be used to further understand individual cases of persons experiencing learning difficulties.

Intellectual Disability

Intellectual disability (formerly called mental retardation) is a nonspecific, heterogeneous disorder that occurs during a person’s early developmental stages (birth to 18 years; Schalock et al., 2007). It is defined in part as involving subaverage general intellectual performance, which in turn is defined as roughly 2 standard deviations or more below average. Of equal importance are difficulties in adaptive behavior, and any assessment of intellectual disability must demonstrate both a low intelligence level (at least roughly 2 standard deviations below the mean) and evidence that the person cannot function independently or deal effectively with day-to-day life problems (American Psychiatric Association [APA], 2013). Poor independent functioning must include at least two adaptive skill areas including communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work. Classification of intellectual disabilities should identify the person’s psychological and emotional strengths and weaknesses, overall physical health, and current environmental placement. The guidelines of the Administration on Intellectual and Developmental Disabilities (AIDD; www.acl.gov/Programs/AIDD) stress that this assessment should lead to a profile that places less emphasis on describing the level of disability (mild, moderate, severe) and more on identifying the types and intensities of supports required by the person. These supports might be intermittent, limited, extensive, or pervasive. Thus, recently there has been a move away from describing the disability in favor of using information about the person to identify how the person’s functioning could be optimized by the best support available for him or her. With appropriate supports, the person’s functioning should be able to improve somewhat over time. In addition, assessment should take into consideration cultural and linguistic diversity and the context of the community environment, as well as balance out the individual’s adaptive limitations with his or her adaptive skills and personal capabilities.

The AAID guidelines emphasize the interaction of the person with the environment. In particular, they encourage any assessment to focus on the level and intensity

of required support with a philosophy of empowering the person. As such, there has been a relative deemphasis on the global IQ score, along with the elimination of person-oriented levels of disability. This does not mean that IQ scores are not important, but there is more of a focus on treatment and community-oriented descriptions. The *DSM-5* (APA, 2013) somewhat mirrors this trend in that it has deemphasized the IQ score, which used to categorize the severity of the disability, and now focuses more heavily on adaptive functioning to determine severity, though the diagnostic criteria are still heavily person-oriented. For most contexts, clinicians should follow the AAID guidelines because they are more useful, more clearly tied to recommendations, represent the most current thinking in the field, and are in accordance with national recommendations. However, many situations may require categorizing diagnosis in line with the *DSM-5* guidelines.

Although intellectual disability (mental retardation) is a heterogeneous disorder, there is consensus that it consists of two general categories. Nonorganic (or familial) retardation is caused by low genetic inheritance, poor environment, and possibly some organic factors. Persons with familial retardation constitute the upper realms of intelligence (50–69) and adaptive functioning among persons with intellectual disabilities and can be educated. Organic retardation is frequently severe ($IQ < 50$) and is more closely associated with neurological dysfunction and genetic impairment. Persons with this disorder typically require more supervision and care but usually can be taught to manage some routine day-to-day activities.

A typical assessment battery for the diagnosis and assessment of intellectual disability includes the WISC-V or other individually administered intelligence tests (K-ABC-II, Stanford-Binet-5), an achievement test (Wide Range Achievement Test-IV, Wechsler Individual Achievement Test-III, Kaufman Test of Educational Achievement [KTEA]), and measures of adaptive functioning (Adaptive Behavior Assessment System/ABAS, AAMD Adaptive Behavior Scale, or Vineland Adaptive Behavior Scales). Further information from interviews, behavioral observations, and medical records is also essential. An important purpose of a test such as the WISC-V is to establish the client's general intelligence level so that it can be placed into the context of other relevant information. When determining the cutoff IQ for diagnosis, the test's range of error should be taken into account. This means that the IQ cutoff criteria are somewhere between 70 and 75. The most difficult subtests for persons with intellectual disabilities are typically Information, Arithmetic, and Vocabulary (primarily the Verbal Comprehension factor; Mueller, Dash, Matheson, & Short, 1984).

Gifted Children

Gifted children are frequently defined as having IQs of 130 or higher. Children who have a single outstanding ability, such as in art, music, or math, are also frequently classified as gifted even though their IQs may not necessarily be above 130. One caution is that portions of the WISC-V and WAIS-IV place considerable emphasis on speeded performance. Thus, a person who is generally gifted, but did not express this giftedness in a rapid manner on the test and achieved a lower Processing Speed Index score, would not appear to be gifted on the Full Scale IQ. This fact may need to be taken into account when making interpretations. However, formal IQ tests may not be particularly good

if a single outstanding ability is used to determine whether a particular child is gifted. Additional assessment strategies for children should include samples of their work, achievement tests, rating forms, or designation by a highly qualified person.

An essential goal of assessing for giftedness is to optimize (rather than normalize) the child's abilities so that a greater likelihood exists that the child will eventually make a significant contribution to society. Assessment of gifted persons typically recommends an appropriate educational placement and provide general guidelines for program planning. IQ, in itself, is in many ways a limited definition of giftedness. Many persons with extremely high IQs do not accomplish anything of significance. A high IQ (or outstanding talent in a specific area) is merely one of a variety of prerequisites, necessary but not sufficient. The interactions of internal motivation, discipline, and environmental opportunities, such as appropriate instruction, are of equal importance.

Caution should also be used when using tests such as the WISC-V to assess gifted persons who demonstrate high creativity. Often highly intelligent people are not particularly creative, a fact that is supported by the low correlation between intelligence tests and creativity (Amabile, 1983). For abilities such as artistic or musical creativity, measures outside IQ testing may prove to be of greater importance. These measures might include a list of creative achievements, nomination by a qualified person, and specific tests of creativity.

SHORT FORMS

Dozens of short forms for the Wechsler intelligence scales have been developed to provide a more time-efficient means of estimating IQ. These short forms reduce administration time by either giving selected subtests or deleting specific items (early easy ones, odd or even items). Although time-efficient, these short forms tend to provide less information about a person's cognitive abilities, produce a wider band of error than a full administration, result in less clinical information, and are often of questionable accuracy when used for intelligence classifications (A. S. Kaufman & Kaufman, 2001; Silverstein, 1990). Indeed, A. S. Kaufman and Kaufman (2001) have recommended not using short forms, especially when well-developed alternative brief intelligence tests are available. However, short forms can serve as screening devices, which are best used when the purpose of evaluation is other than for intellectual assessment and there are no suspicions that cognitive functioning is significantly outside of the norm. The results can be used either as a rough indicator of intelligence or as a basis for determining whether a more complete cognitive assessment is necessary. None of the short forms should be confused with a full intellectual assessment or even with a valid indicator of IQ. For this reason, it is important to clearly specify on the report that the indicated IQ is an estimate (indicate as *Est* next to the IQ score) and that a "brief" WAIS-IV/WISC-V was given. If this is not specified, the IQ derived from the short form may be confused with a full administration, and later decisions may be based incorrectly on the misleadingly described results.

The basic requirement for any short form is a minimum correlation of .90 with the full administration. Even at the .90 level, the band of error is considerably wider than for an IQ derived from a full administration. Calculations indicate that at a .90

correlation, two-thirds of the IQs fall within 9 points of a person's actual IQ and a full one-third are 10 or more points away from the actual IQ (L. Schwartz & Levitt, 1960). In addition to these psychometric considerations, short forms might be selected based on the type of clinical information needed or special client characteristics (e.g., disabled, non-English-speaking background).

Many clinicians calculate short-form IQs by prorating the subtest scores—that is, by calculating the mean subtest score for the subtests that were given. This mean can then be multiplied by the total number of core subtests (10) to derive an estimated Full Scale sum of scaled scores. Once this estimate of sum of scaled scores has been determined, relevant tables in the manual(s) can be consulted to determine the estimated IQs. The *WAIS-IV Administration and Scoring Manual* provides some prorating for calculating the Full Scale IQ when nine subtests have been given (see Table A.9, p. 227) or for calculating the Verbal Comprehension Index and Perceptual Reasoning Index when two subtests have been given (see Table A.8, p. 226). Similarly, the *WISC-V Administration and Scoring Manual* offers a prorating option for when only six of the seven necessary subtests have been given (see Table A.8, p. 336). Unfortunately, prorating may produce error by failing to consider the relative reliabilities of the different subtests that were used. To counter this problem for the WISC-IV, Sattler (2008) provided conversion tables for combinations of two, three, and five subtest short forms. Sattler also provided a formula for obtaining deviation IQs from short forms. Similar resources for the WAIS-IV and WISC-V are not yet available.

Wechsler Abbreviated Measure of Intelligence

The Psychological Corporation developed the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999) as a means of providing clinicians and researchers with a short, reliable measure of intelligence linked to the WAIS-III (and WISC-III). The WASI-II (Wechsler, 2011) came out in 2011 as an update, primarily to improve psychometric properties and link it more closely with the WAIS-IV and WISC-IV. The WASI includes four subtests (Vocabulary, Similarities, Block Design, and Matrix Reasoning), which have a similar format and similar content as the WAIS-IV and WISC-IV/WISC-V subtests with the same names. The selection of these subtests was based in part on high loadings on *g*, along with evidence suggesting bilateral hemispheric activation on most complex cognitive tasks (Springer & Deutsch, 1998). The WASI yields both Verbal and Performance IQs as well as a Full Scale IQ. The WASI-II was nationally standardized using a population ranging between ages 6 and 89. Because the subtests were linked to the longer Wechsler intelligence scales, the WASI provides reliable estimates of full WAIS-IV and WISC-IV (and likely WISC-V) IQs. Administration time can be reduced even further by using a two-subtest form (Vocabulary and Matrix Reasoning), which takes approximately 15 minutes but yields only a Full Scale IQ estimate. If it is determined from the two- or four-subtest administration of the WASI-II that a fuller administration of a cognitive battery is necessary, the four subtests on the WASI-II can substitute for the subtests of the same name on the WAIS-IV and WISC-IV (and likely the WISC-V, though this has not yet been established). If a brief, Wechsler-based version of intelligence is desired, it is strongly recommended that the WASI-II be used in preference to the short forms described next.

Best Two- and Three-Subtest Short Forms

One of the most popular two-subtest short forms uses Vocabulary and Block Design. Administration time is approximately 20 minutes, and correlations with the full-administration Full Scale IQ are generally in the .90 range (Sattler, 2001, 2008). In two-thirds of the cases, IQs fall within 7 points of a person's actual IQ, and one-third of the scores have an error of 8 points or greater. Conceptually, Vocabulary and Block Design are good tests to use because they are both good measures of *g*, are quite stable, and represent a sample subtest from both the Verbal Comprehension and the Perceptual Reasoning/Visual Spatial indexes. However, research with the WAIS-R suggested it might potentially underestimate the IQs of African Americans because these two subtests are typically their lowest scores (Kaufman et al., 1988). Furthermore, persons with high IQs are likely to have a greater margin of error when short forms are used to estimate their IQs because of the greater degree of subtest scatter among this subgroup (Matarazzo, Daniel, Prifitera, & Herman, 1988). If examiners wish to add a third subtest, the inclusion of Similarities, Information, Comprehension, and Picture Completion have each been found to increase correlations into the low .90s (Sattler, 2001, 2008).

Best Four-Subtest Short Forms

Short forms using any four combinations of Vocabulary, Similarities, Symbol Search, Arithmetic, Coding, Block Design, Arithmetic, Matrix Reasoning, Digit Span, and Information are likely to produce correlations with the Full Scale IQ in the mid-.90s (Sattler, 2008). Decisions on which subtests to include may depend on the type of information that is required. For example, including Vocabulary provides an indication of a person's verbal abilities and is the best predictor of Full Scale IQ. However, for ethnic minorities or those who might have English as their second language, a focus on nonverbal subtests may provide a better estimate of their abilities (e.g., Block Design, Matrix Reasoning) than verbal subtests like Vocabulary. In order to screen for impaired cognitive functioning, "brain-sensitive" tests such as Coding and Symbol Search might be important to include. If more verbally oriented subtests are given (e.g., Vocabulary, Information), it may overestimate their level of functioning.

Seven-Subtest Short Forms

One strategy is to delete the most time-consuming subtests and give as many of the shorter subtests as possible. For example, J. J. Ryan and Ward (1999) developed a WAIS-III seven-subtest short form (Information, Digit Span, Arithmetic, Similarities, Picture Completion, Block Design, Coding), which takes 35 minutes to administer. A slight variation from this short form is to substitute Matrix Reasoning for Block Design. This combination is likely to also be time efficient for the WAIS-IV, but the administration time and reliability and validity data have not been calculated. However, data on the WAIS-III indicated that estimated Full Scale IQ scores were nearly as reliable as for full-administration IQs with the average Full Scale SEM being 2.80 (and 2.72 for the version with Matrix Reasoning) versus 2.58 for the full WAIS-III

Full Scale IQ (J. J. Ryan & Ward, 1999). Correlations between the J. J. Ryan and Ward (1999) seven-subtest short form and a full administration were .98 for the Full Scale IQ, .97 for the Verbal IQ, and .95 for the Performance IQ (.96 using Matrix Reasoning). Thus, the psychometric properties of the seven-subtest short form are excellent. For the WISC-V, a seven-subtest short form is included as only seven subtests are required to calculate the Full Scale IQ. In addition, this seven-subtest version will give index scores for Verbal Comprehension and Fluid Reasoning but not for Visual Spatial, Working Memory, or Processing Speed.

Additional Short Forms (Satz-Mogel/Yudin and Modified Formats)

An alternative to administering various combinations of subtests is to use every subtest but limit the number of items from each one. The most frequently used variation is the Satz and Mogel (1962) approach, which was originally developed for the WAIS but can also be used for the WAIS-IV. The procedure is to administer every third item for Information and Vocabulary and multiply the scores by 3 to obtain the raw scores. Only odd items would be administered for Similarities, Arithmetic, Block Design, and Visual Puzzles, and each score is multiplied by 2 to obtain the respective scaled scores. Full administrations would be given for Digit Span, Coding, Matrix Reasoning, and Symbol Search. The entire procedure for the WAIS-IV took approximately 40 minutes, and the derived IQs had correlations similar to the best four-subtest variations. A distinct advantage over four-subtest variations is that the Satz-Mogel approach samples a wider range of areas. This is likely to increase the stability of scores over a wider variety of populations and allows clinicians to develop inferences over a larger number of behaviors. Research with the WAIS-III indicated that IQs derived from the Satz-Mogel usually did not vary more than 6 points when compared with the full administration (J. J. Ryan, Lopez, & Werth, 1999). In addition, a full 86% of the clients had the same IQ classifications. A caution is that, although a score is provided for each subtest, it is inappropriate to attempt a profile analysis because the individual subtests are not sufficiently reliable (J. J. Ryan et al., 1999). Given that it would take only an additional 20 minutes to administer the entire battery, the entire administration seems a worthwhile time investment.

A WISC-V equivalent of the Satz-Mogel approach would have the same advantages and disadvantages and follow a nearly identical procedure. However, Arithmetic and Information would not be given because they are optional subtests; Coding would be given in its entirety. In addition, every other item would be given for the core WISC-V Picture Concepts and Figure Weights subtests, and Figure Span would need to be given in its entirety.

A final approach is the elimination of early, easy items on each of the subtests. This is most appropriate for relatively bright subjects but should be used cautiously with persons of below-average intelligence. Cella (1984) provided WAIS-R guidelines for the number of items to be omitted based on a subject's performance on the Information subtest. Research on the WAIS-IV and WISC-V using this format has not been conducted. However, research with the WAIS-R has shown an almost exact correlation (.99) with a full administration, yet this approach can reduce the total administration time by 25%. Despite this high correlation, some caution should be exercised regarding

Cella's Modified Format and the Satz-Mogel approaches. First, lowered internal consistency is likely to reduce subtest reliabilities sufficiently to render profile analysis questionable. Second, examinees are disadvantaged because they are not able to have as many previous subtest items to practice on (as items are skipped) before more difficult items are administered. The result may be that the norms for the full administration may not necessarily apply to the shortened versions. Again, the slight reduction in time is probably not worth the loss of psychometric quality.

RECOMMENDED READING

- Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Boston, MA: Allyn & Bacon.
- Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Hoboken, NJ: Wiley.
- Sattler, J. M. (2008). *Assessment of children: Cognitive functions* (5th ed.). San Diego, CA: Author.
- Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. San Diego, CA: Sattler.
- Weiss, L.G., Saklofske, D., Holdnack, J., & Prifitera, A. (2015). *WISC-V assessment and interpretation*. St. Louis, MO: Elsevier.

WECHSLER MEMORY SCALES

The Wechsler Memory Scales (WMS) are individually administered, composite batteries designed to enable the user to better understand various components of a client's memory. Now in its fourth edition (WMS-IV), it has been conormed with the WAIS-IV. Another major feature is that it provides a full range of memory functioning and has been carefully designed according to current theories of memory. As a result of these features, it typically is considered to be a core component of any thorough cognitive assessment (Rabin, Barr, & Burton, 2005), which is reflected in its being ranked as the ninth most frequently used test by clinical psychologists (and third most frequently used by neuropsychologists; Camara, Nathan, & Puente, 2000).

Memory complaints are extremely prevalent among client populations. These complaints are associated with anxiety, schizophrenia, depression, head injuries, stroke, learning disabilities, and neurotoxic exposure, among many other problems. For example, the impact of alcohol and other drugs on memory might need to be evaluated carefully. In occupational contexts, the clinician similarly might need to evaluate the impact of exposure to industrial agents (e.g., lead, mercury, organic solvents) that can potentially result in impaired memory function. The increasingly aging population means that distinguishing normal memory loss from the expression of dementia will become progressively more important. One crucial differential diagnosis is to distinguish between what used to be termed *pseudodementia* (memory impairment resulting from depression) and true dementia, such as Alzheimer's disease. As various drugs are developed to treat cognitive difficulties, it will become increasingly important for clinicians to monitor client improvement with a particular emphasis on memory functioning. The array of memory symptoms suggests a developmental perspective: Children are most likely to experience memory complaints related to learning disabilities and attentional problems; adults typically experience difficulties because of neurotoxic exposure or head injuries; and older populations have memory problems related to conditions causing dementia.

Many early conceptualizations of memory considered it a unitary process. From a practical assessment perspective, it was not necessary to have a composite battery that assessed various components of memory. In contrast, more recent conceptualizations consider memory to have various components (see Lezak, Howieson, Bigler, & Tranel, 2012), with a cognitive sequence. First, an individual needs to attend to presented material, so attention and concentration play a role in memory. This attention and subsequent processing can involve various sensory components, such as visual and auditory modes of processing. Active engagement is also important; thus, "working memory" was conceptualized as containing an executive component that initially monitors and evaluates information. Researchers also distinguish between short-term

and long-term memory (sometimes described as primary and secondary memory storage, respectively), often discussed as an encoding or storage process. Finally, memory retrieval requires yet another mental process. Within this cognitive sequence, a further well-supported distinction is between memory that is conscious and reflected in verbal reports of facts, events, and experiences (*declarative, explicit, or episodic memory*) versus memory that is more unconscious and measured implicitly by changes in performance (*procedural or implicit memory*).

HISTORY AND DEVELOPMENT

In some ways, the development of the Wechsler Memory Scales (WMS) has paralleled the development of knowledge on memory. Each of the four editions of the scales has increasingly incorporated advances in the theoretical understanding of memory. The original Wechsler Memory Scale (Wechsler, 1945) reflected earlier, nonspecific conceptualizations of memory. It was composed of brief procedures on memory for number sequences, recalling a story, simple visual designs, and paired words. The advantage of using a variety of procedures was that a client might have intact memory for visual information but not auditory information or vice versa. Despite the fact that the early WMS procedures could be *logically* divided into visuospatial versus auditory tasks, the overall scoring was a composite Memory Quotient that, similar to the Wechsler intelligence scale Intelligence Quotients (IQs), had a mean of 100 and a standard deviation of 15. This was extremely valuable information for clinicians because they could easily compare a client's IQ with his or her Memory Quotient and further investigate any large discrepancy. The Wechsler Memory Scale was also quite popular as it was a relatively brief procedure, typically taking about 15 minutes to complete. Because retesting a client would be likely to result in practice effects, it had the further advantage of having a parallel form. As a result of these advantages, it became a widely adopted procedure among clinicians.

Despite the fact that the Wechsler Memory Scale had surprising longevity (a formal new version did not become available until 1987, a 42-year interval), it had several limitations. First, it included unsophisticated methods of scoring the various procedures. In addition, the algorithms to determine the Memory Quotient were overly simple as they did not consider a sufficient number of client variables. The norms were derived from a small sample of 200 patients between the ages of 25 and 50 at Bellevue Hospital in New York. Scores for either older or younger persons were extrapolated from this sample but were not based on actual participants. In addition, the alternate form was rarely used, and the research supporting it was quite limited. Finally, it did not reflect advances in knowledge related to memory processes.

One early attempt to correct for the deficiencies of the Wechsler Memory Scale was Russell's (1975, 1988) adaptation, in which he administered two of the subtests (Logical Memory and Visual Reproduction) in an immediate format combined with a delay of 30 minutes. This method allowed comparisons to be made between short-term and long-term memory. Research on Russell's WMS supported the predicted difference between left-hemisphere (relatively lowered auditory recall based on Logical Memory) and right-hemisphere (relatively lowered visual reproduction

based on Visual Reproduction) lesions. Despite these advantages, the psychometrics were weak, and the test was poorly standardized. Unfortunately, it was titled the Wechsler Memory Scale–Revised (WMS-R), which could create confusion because The Psychological Corporation developed a full revision of the WMS that was also titled the Wechsler Memory Scales–Revised (WMS-R). Subsequent publications have attempted to clarify the two versions by referring to them as either Russell’s WMS-R or the WMS-R.

The 1987 revision (Wechsler Memory Scales–Revised [WMS-R]) was a significant improvement over the WMS. It had age-related norms for nine different age groups ranging from 16 to 17 years old for the youngest group and 70 to 74 years for the oldest group. The standardization sample was composed of 316 persons, who had demographic characteristics that closely approximated 1980 census data. There were approximately 50 subjects in each of the age groups. Whereas the Wechsler Memory Scale had only one composite Memory Quotient, the WMS-R had 12 subtests from which five composite scores could be derived: General Memory, Attention-Concentration, Verbal Memory, Visual Memory, and Delayed Recall. Each of the index scores had a mean of 100 and a standard deviation of 15. This division into index scores is consistent with theories that have divided memory into short-term versus long-term (note the Delayed Recall used to assess long-term memory) and verbal/auditory versus visual (note the Verbal Memory and Visual Memory indexes).

Reliability of the WMS-R was generally low to adequate (internal consistency ranged from .44 to .88, and test-retest reliabilities ranged from .51 to .60). The index standard error of measure ranged from a high of 8.47 for the Visual Memory Index to a low of 4.86 for the Attention-Concentration Index (Wechsler, 1987). As with studies on reliability, the validity of the WMS-R was good to adequate. Factor-analytic studies supported either a two- (Bornstein & Chelune, 1988; Roid, Prifitera, & Ledbetter, 1988; Wechsler, 1987) or three-factor solution (Bornstein & Chelune, 1988). A wide range of studies supported the ability of the WMS-R to distinguish between normal and clinical groups (K. A. Hawkins, Sullivan, & Choi, 1997; Reid & Kelly, 1993; Wechsler, 1987), distinguish the relative severity of deficits based on subjective complaints (Gass & Apple, 1997), provide an index that related to client ratings of level of everyday memory (Reid & Kelly, 1993), and predict the degree of brain atrophy (Gale, Johnson, Bigler, & Blatter, 1995). In addition, the Attention-Concentration Index was found to be one of the most sensitive available measures in identifying cognitive impairment (M. Schmidt, Trueblood, Merwin, & Durham, 1994). Despite a conceptual basis for believing that visual and verbal memory would relate to brain laterality of deficits, research on this theory has produced inconsistent results (Chelune & Bornstein, 1988; Loring, Lee, Martin, & Meador, 1989).

The WMS-R had clear advantages over the WMS, as it had a far better normative base, was validated on a diverse sample, had quite extensive studies performed on it, and divided memory into various indexes, thereby allowing clinicians the possibility of measuring various aspects of memory. Nevertheless, its weaknesses resulted in a revision within a relatively short period. One of the most serious limitations of the WMS-R was the relatively low reliabilities of the subtests and indexes (Elwood, 1991), which likely significantly reduced the accuracy of measurements. In addition, the different indexes were probably not very good measures of specific components of memory.

This is not to say they were not sensitive to both *general* cognitive impairment and the *degree* of that impairment. However, the specific nature of the impairment could not be determined accurately by referring to the specific indexes, despite the fact that the index names suggested that this differentiation could be made. Finally, current theories of memory were not used in the design of the WMS-R (Lichtenberger, Kaufman, & Lai, 2002).

The Wechsler Memory Scale–Third Edition (WMS-III) was published just 10 years after the release of the WMS-R (Wechsler, 1997b, 2002a, 1997b, 2002b). The new revision was designed not merely as a face-lift of the WMS-R but rather a “state of the art assessment instrument that comprehensively address[ed] the complexity of brain/behavior relationships involved in learning and memory” (Edith Kaplan, foreword to WMS-III manual, 2002b, p. iii). To accomplish this goal, new subtests were added, scoring procedures were made more sophisticated, stimulus materials were changed, and new index configurations were developed. This resulted in six primary and five optional subtests that were organized into index scores. Whereas the manual stated that it is possible to administer the six primary subtests in 30 to 35 minutes, research with a clinical population indicated that it took 42 minutes to administer the 11 primary subtests (see Axelrod, 2001). An abbreviated version was published in 2002 that reduced the administration time to 15 to 20 minutes (WMS-III Abbreviated; Wechsler 2002b).

One of the most important aspects of the WMS-III is that it was developed simultaneously with the WAIS-III. This enabled the two tests not only to share two subtests but also to be conormed. The normative sample consisted of 1,250 adults ranging between 16 and 89 years. Instead of nine groups, as in the WMS-R, the WMS-III had 13 different groups. These groups not only had more subjects (100 for each of the first 11 groups of the WMS-III versus 50 in each group for the WMS-R) but also extended to a far higher age range (89 for the WMS-III versus 74 for the WMS-R). This extension is appropriate because one of the more important functions of memory assessment is to evaluate older adults.

The WMS-III had better reliability than its predecessor. The *WAIS-III/WMS-III Technical Manual* indicated that internal consistency for the primary subtest scores ranged between .74 and .93 for all age groups. As would be expected, the primary indexes had better internal consistencies, at .82 or higher. Test-retest reliabilities for all age groups over a 2- to 12-week interval mostly ranged between .62 and .82 for the individual subtests and between .75 and .88 for the indexes. The *Technical Manual* stated that even those subtests requiring the most judgment (Logical Memory I and II, Family Pictures I and II, Visual Reproduction I and II) had interscorer reliabilities above .90.

Factor analyses reported in the 1997 *WMS-III Technical Manual* concluded that a three-factor model composed of Working Memory, Visual Memory, and Auditory Memory most closely fit the data. In contrast, the 2002 *Technical Manual* found that a five-factor model composed of Working Memory, Auditory Immediate Memory, Visual Immediate Memory, Auditory Delayed Memory, and Visual Delayed Memory best fit the age groups from 30 to 64 and 65 to 89. For ages 30 to 89, this factor structure closely corresponded to five of the eight index scores. The change in factor structure between younger and older age groups is also consistent with findings that the

components of memory become more clearly distinguishable (“dissociated”) with age (Dolman, Roy, Dimeck, & Hall, 2000). Thus, the individual index scores might become more meaningful with older populations. However, it should be noted that for most populations, there was a correlation of .98 between Immediate Memory and General Memory (K. A. Hawkins, 1998; Weiss & Price, 2002). This finding suggests that much of the time, these indexes were redundant. Other research has found a four-factor model (Auditory, Visual, Working Memory, Learning; Price, Tulsky, Millis, & Weiss, 2002) and a two-factor model (Wilde et al., 2003).

There was ample evidence that the WMS-III effectively differentiated between clinical and normal populations. Various clinical groups (including Alzheimer’s, Huntington’s, and Parkinson’s disease; multiple sclerosis; chronic alcohol abuse; temporal lobe epilepsy; schizophrenia) consistently scored lower than the standardization sample (Fisher, Ledbetter, Cohen, Marmor, & Tulsky, 2000; K. A. Hawkins, 1998; Psychological Corporation, 1997). For example, patients with Korsakoff’s disease typically have severe difficulties with encoding and storing new information, but their attention and working memory are normal. This finding was reflected on the WMS-III index performances, with Working Memory in the normal range but all other index scores in the impaired range (Psychological Corporation, 1997). In addition, patients with mild Alzheimer’s disease scored in the 60 to 71 range for most of the primary indexes, except for a mean score of 80 for Working Memory (Psychological Corporation, 1997). Fisher and colleagues found that patients with moderate to severe traumatic brain injury scored low on all indexes. This finding is also consistent with the finding that clinician ratings of the severity of brain injury were accurately reflected on scores on the WMS-III (Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999). The previous sampling of studies indicates that many of the predicted theoretical and clinical patterns of performance were supported.

Despite the relative success of the WMS-III, it still had a number of limitations. Some of the most important of these limitations were the equivocal factor structure, long testing time for older adults, subtest overlap with the WAIS-III, and specific problems with some of the subtests (Faces, Family Pictures, Verbal Paired Associates). To correct for these limitations in the WMS-III and refine memory assessment further, the WMS-IV was developed and published in 2009 (Pearson, 2009c, 2009d). To counteract the WMS-III limitations, the WMS-IV organized the indexes according to a clear factor structure, tried to reduce the administration time (especially for older adults), eliminated the subtest overlap with the WAIS-IV (deleting Digit Span and Letter-Number Sequencing), and eliminated some of the problematic subtests (Faces, Family Pictures, Spatial Span, Word Lists). Additional innovations included increasing the score ranges (ceiling and floor), focusing on *visual* working memory tasks (versus the WAIS-IV emphasis on *auditory* working memory tasks), adding a new subtest (Spatial Addition), clarifying/simplifying some of the scoring procedures (Visual Reproduction), and modifying some of the WMS-III subtests (Logical Memory, Verbal Paired Associates, Symbol Span, Designs).

These efforts resulted in six subtests plus an additional optional Brief Cognitive Screen (see Table 6.1). Most of the subtests are administered a first time, then readministered following a 20- to 30-minute delay. Theoretical and factor analyses clustered the subtests according to five indexes (see Table 6.2). This is in contrast to the seven indexes

Table 6.1 WMS-IV Subtest Names, Descriptions, and Abilities Measured

Subtest title	Description
Brief Cognitive Status (Optional)	Basic tasks include orientation (year, date, etc.), mental control (counting backward), drawing a clock, recall of objects that had been named previously, inhibition of responses, and verbal production. A total score is derived to provide an estimate of any major cognitive impairment (<i>gross cognitive impairment</i>)*.
Logical Memory (Ages 16–90)	I*: Two short stories are read, and examinees are requested to repeat as many details as possible. Older adults (65–90) are presented with only one story presented twice (<i>short-term auditory-verbal memory</i>). II: Examinees are again asked to recall as many of the details as possible (<i>long term auditory-verbal recall</i>). They are then asked yes/no questions about details in the stories (<i>long-term auditory-verbal recognition</i>).
Verbal Paired Associates (Ages 16–90)	I: A list of pairs of words are read (e.g., “dark ... light”). The first word of the pair is read again, and examinees are asked to remember what the second word of the pair is (e.g., (dark ... ?”)) (<i>short-term auditory learning</i>). II: Examinees are again read the first word in the list and requested to recall the correct paired word (e.g., “light ... ?”) (<i>long-term auditory memory</i>). They are then read a list of paired words and asked which pairs were/were not read to them in condition I (<i>long-term auditory recognition</i>). They are then asked (optional task) to say as many of the word pairs as they can recall (<i>long-term auditory recall information</i>).
Designs (Ages 16–69)	I: Examinees are shown a series of designs placed on a grid (10 seconds). The grid is removed, and a new grid is presented along with a set of designs. Examinees must then identify where on the grid the original designs belong (<i>short-term spatial memory</i>). II: Examinees are shown designs and grids and asked to reproduce the original placement of the designs on the grid (<i>long-term visuospatial memory</i>). They are then shown grids with designs on them and asked to recognize which designs are the same as in the immediate (I) condition (<i>long-term visual recognition</i>).
Visual Reproduction (Ages 16–90)	I: Examinees are shown five designs for 10 seconds. They must then draw the designs from memory (<i>short-term visual memory</i>). II: Examinees are requested to draw the original designs from memory (free recall task) (<i>long-term visual memory</i>). Next they are asked to identify which of six designs on a page is the same as the design shown in condition I (<i>long-term visual recall</i>). Finally (optional task), examinees are shown the original designs and requested to draw them (copy phase; <i>visuospatial construction</i>).

Table 6.1 (Continued)

Subtest title	Description
Spatial Addition (Ages 16–90)	Examinees are shown two grids with blue and red circles. They are then asked to add or subtract the location of the circles but are guided by a set of rules (<i>visual-spatial working memory</i>).
Symbol Span (Ages 16–90)	Examinees are first shown a page with a series of abstract symbols. They are then shown a different array of symbols and are asked to identify the correct order that they were presented on the original page (<i>visual-sequencing working memory</i>).

*I indicates that the procedure was administered and the client's memory for the activity was assessed immediately afterward ("immediate" condition). In contrast, II indicates that a variation on the original procedure (I) occurred 20 to 30 minutes later ("delayed" condition).

**The phrases in italics indicate the type of memory function that is measured by the subtest.

Source: Adapted from Table 1.1 from the *WMS-IV Administration and Scoring Manual* by Pearson, 2009, Pearson, Inc.

Table 6.2 WMS-IV: Adult Battery (Ages 16–69) Indexes, Primary Subtests

Index	Subtests used to calculate index
Auditory Memory	Logical Memory I, Verbal Paired Associates I, Logical Memory II, Verbal Paired Associates II
Immediate Memory	Logical Memory I, Verbal Paired Associates I, Designs I, Visual Reproduction I
Delayed Memory	Logical Memory II, Verbal Paired Associates II, Designs II, Visual Reproduction II
Visual Memory	Designs I, Visual Reproduction I, Designs II, Visual Reproduction II
Visual Working Memory	Spatial Addition, Symbol Span

for the WMS-III. Another major change was the development of one battery for adults (*WMS-IV: Adult Battery*; ages 16–69) and a slightly modified battery for older adults (*WMS-IV: Older Adult Battery*; ages 65–90, see Table 6.3). While an advantage of the Older Adult Battery is that administration time is shorter, only four indexes can be calculated. In order to assist with interpretation, a series of "contrast scores" were included that determine whether differences between subtests were large enough to be interpretable. For example, a Visual Memory Index that was significantly higher than a client's Auditory Memory Index suggests that visual ability is a relative strength. Interpretation of these differences could assist with diagnosis and guide the development of treatment recommendations.

The standardization sample for the WMS-IV was representative of the 2005 U.S. census of persons between the ages of 16 and 90. As such, the sample reflects the U.S. population based on age, sex, race/ethnicity, education level, and geographic region. A total of 1,400 examinees were included with 100 in each of 14 age bands. A wide variety of exclusion criteria was employed to make sure that inappropriate persons

Table 6.3 WMS-IV: Older Adult Battery (Ages 65–90) Indexes and Primary Subtests

Index	Subtests used to calculate index
Auditory Memory	Logical Memory I, Verbal Paired Associates I, Logical Memory II, Verbal Paired Associates II
Immediate Memory	Logical Memory I, Verbal Paired Associates I, Visual Reproduction I
Delayed Memory	Logical Memory II, Verbal Paired Associates II, Visual Reproduction II
Visual Memory	Visual Reproduction I, Visual Reproduction II

were not included (e.g., persons with dementia, psychosis, or medication that might impair their performance). The WMS-IV was conformed with the WAIS-IV, thereby making comparisons between the two instruments more appropriate.

RELIABILITY AND VALIDITY

The WMS-IV has good to excellent reliability. Subtest internal consistency among the normative groups was highest for Visual Reproduction II (.97) and Verbal Paired Associates (.94) and lowest for Visual Reproduction (.74) and Verbal Paired Associates (.76; Pearson, 2009d). Internal consistency among a wide variety of clinical groups (e.g., Alzheimer's disease, traumatic brain injury) was even higher. As would be expected, the index score internal consistencies were all excellent, ranging from a high of .98 for Visual Memory (standard error of measurement [SEM] = 3.04) to a low of .93 for Visual Working Memory (SEM = 3.71; Pearson, 2009d). Subtest test-retest reliabilities over a 14- to 84-day interval ($M = 23$ days) ranged between .77 for Spatial Addition and a low of .59 for Designs I (Spatial scoring category). Since the index scores have a larger number of items/subtests, it would be expected that their test-retest reliabilities would be higher than for the individual subtests. This finding was supported in that the index test-retest reliabilities ranged from a high of .83 (Auditory Memory and Visual Working Memory) to a low of .81 (Visual Memory and Immediate Memory).

Extensive and quite supportive evidence for the validity of the WMS-IV is presented in the *WMS-IV Technical and Interpretive Manual* (Pearson, 2009d), which can be organized according to content validity, correlations among the WMS-IV subtests/indexes themselves, factor analyses, correlations with other measures, and differentiation between normal populations and special groups (e.g., traumatic brain injury, intellectual disabilities, Alzheimer's disease). Content validity was based on a combination of research on previous versions of the Wechsler Memory Scales, expert review, client feedback, and research on the cognitive processes clients underwent when responding to the test items. Based on this information, the test items were modified and eventually evolved into those on the current WMS-IV. As a result, considerable efforts have been made to develop and refine their content validity.

The *WMS-IV Technical and Interpretive Manual* (Wechsler, 2009b) presents additional information related to correlations among the various subtest/index scores. Ideally, subtests/indexes on a test such as the WMS-IV would be expected to have

positive correlations with similar tests (convergent validity) and low or nonexistent correlations with tests that do not seem similar (discriminant validity). Among the WMS-IV subtests/indexes themselves, they all would be expected to have at least some correlation with each other since each of the subtests measures, to some extent, the common variable of memory. This was indeed the case. In addition, measures of similar abilities would be expected to have somewhat higher correlations than measures of dissimilar abilities. As an example, the verbal subtest Logical Memory I was found to have a moderate, positive correlation with Verbal Paired Associates (.44). In contrast, a lower correlation was found between Verbal Paired Associates (a verbal subtest) and the more visual subtest Spatial Addition (.31). However, the correlation between the Delayed Memory and the Immediate Memory indexes was quite high, .87. This finding is high enough to suggest that the subtests may be measuring quite similar constructs.

In contrast, the Auditory Memory and Visual Memory indexes were moderately correlated (.64), suggesting that the auditory and visual memory components of the WMS-IV are adequately differentiated.

Previous factor analyses of earlier editions of the Wechsler Memory Scales resulted in inconsistent findings, which created considerable debate regarding the true structure of the instruments and called into question the accuracy of some WMS-R/WMS-III index groupings. As a result, the WMS-IV closely adhered to a factor-analytically supported three-factor model comprised of Auditory Memory, Visual Memory, and Working Memory. Immediate Memory and Delayed Memory indexes were also included even though they were highly correlated. The rationale for including these last two indexes was that they have been found to be clinically useful constructs among some clinical groups where short-term acquisition of memory occurs but then decays over time (Millis, Malina, Bowers, & Ricker, 1999).

A wide number of correlations with other similar measures support the concurrent validity of the WMS-IV. For example, the correlation between the California Verbal Learning Test—II (CVLT—II) learning trials 1–5 and the WMS-IV Auditory Memory Index was .63 (Pearson, 2009d). Similarly, correlations between the Children's Memory Scales (CMS; for 16-year-olds) and the WMS-IV ranged between a high of .74 (for Immediate Memory—CMS General Memory) to a low of .25 (for Auditory Memory—CMS Visual Immediate). In addition to correlations with specific measures of memory, the *WMS-IV Technical and Interpretive Manual* (Wechsler, 2009b) provides numerous correlations with more general ability measures. For example, index correlations with the WAIS-IV range from .71 (between Full Scale IQ and Visual Working Memory) to a low of .40 (between Auditory Memory and Processing Speed). As would be expected, the highest subtest correlations were between WMS-IV and WAIS-IV spatial measures (i.e., Spatial Addition-Block Design, $r = .51$). One of the most important functions of a psychological test is to make accurate predictions related to everyday behavior. In support of this, the WMS-IV demonstrated positive correlations with a measure of independent living (the Independent Living Scales Full Scale and WMS-IV Immediate Memory Index, $r = .51$). Finally, the WMS-IV moderately correlated with measures of achievement on the Wechsler Individual Achievement Test—II (e.g., WMS-IV Visual Working Memory Index and WIAT—II Mathematics, $r = .77$).

The *WMS-IV Technical and Interpretive Manual* (Wechsler, 2009b) provides results for a wide variety of special groups. A sampling of some of these results is provided here, including for persons with mild intellectual disabilities, Alzheimer's disease, traumatic brain injury, and schizophrenia. Again, data have been highly supportive of the validity of the WMS-IV. As would be expected, scores for persons with moderate intellectual disabilities were low, ranging between a low of 49 on Immediate Memory to a high of 54 for Auditory Memory. Patients in the early stages of Alzheimer's disease typically report memory as being their primary complaint. As a result, it would be expected that their WMS-IV scores would be lower than their WAIS-IV scores. This expectation was supported; mean WMS-IV scores for patients with mild Alzheimer's disease ranged between 64 for Delayed Memory and 72 for Immediate Memory. In contrast, their mean WAIS-IV General Ability Index was a significantly higher 87. The subtests that were the most difficult for patients with mild Alzheimer's were Logical Memory (Scaled score $M = 2.20$) and Verbal Paired Associates (Scaled score $M = 2.05$). It was found that patients with mild to severe traumatic brain injury had WMS-IV scores that were significantly lower than the standardization group, ranging from a high of 86 for Visual Working Memory to a low of 78 for Delayed Memory. Individuals with schizophrenia also had lower WMS-IV scores, ranging between a high of 82 for Visual Memory to a low of 77 for Immediate Memory.

Research reported in the *WMS-IV Technical and Interpretive Manual* (Wechsler, 2009b) amply support differentiating between normal and clinical groups. However, what is particularly crucial for the practicing clinician is to determine whether the individual indexes can accurately measure subcomponents of memory. Factor-analytic studies and determining whether patterns of scores match theories of memory are particularly important. As noted previously, the WMS-IV indexes were carefully organized according to the results of factor analysis. A further area of investigation is to see whether expected index patterns occur among specific types of clinical populations. For example, it would be hoped that the WMS-IV visual and auditory index scores would reliably differentiate patients with right-hemisphere brain damage (lower visual memory scores) from those with left-hemisphere brain damage (lower verbal/auditory scores). Some support for this was found among patients with right temporal lobe epilepsy, who had lower Visual Memory scores ($M = 86$) compared to scores on Auditory Memory ($M = 95$). Moreover, patients with left temporal lobe epilepsy had, as expected, lower Auditory Memory scores ($M = 78$) than Visual Memory scores ($M = 98$). Additional future research will no doubt investigate the extent to which the WMS-IV can differentiate discrete components of memory.

ASSETS AND LIMITATIONS

The WMS-IV is generally an excellent instrument capable of measuring a wide range of memory functioning. It has been based on theoretical research into the processes of memory, it has excellent standardization, and most research indicates solid empirical support. Subtests found to be problematic on the WMS-III were eliminated or modified, and a new subtest was added. There are only five indexes (four for the Older

Adult Battery), and these are consistent with theories of memory, have generally good empirical support, and should make interpretation easier than the seven indexes developed for the WMS-III. The WMS-IV has been conformed with the WAIS-IV, which allows practitioners to make realistic comparisons between performance on the two instruments. In addition, the shorter format for older adults (ages 65–90) has the advantage of making the test more user friendly for this population. The WMS-IV is clearly an improvement over previous editions.

The scoring and administration of the WMS-IV is, for the most part, clearly described in the manual. The artwork is also clear, as is the Record Form. However, Logical Memory does not present guidelines regarding the speed at which the stories should be read. It also does not have guidelines for intonations, pauses, or inflections. Examiner variation in each of these areas may, therefore, result in the potential for error. For the WMS-III, Lichtenberger, Kaufman, and Lai (2002) recommended that the test developers introduce an audiotaped administration. This might be considered for the WMS-IV as well. A further issue with both Logical Memory I and II is its high degree of cultural loading; therefore, persons whose first language is not English may be disadvantaged on this subtest.

The original WMS had the advantage of taking only 15 minutes to administer. The WMS-III increased the administration time to an average of 42 minutes, but it may have actually taken up to 100 minutes for some clinical populations (Lichtenberger et al., 2002). Since the WMS-IV reduced the number of subtests and resulting indexes, it would seem reasonable that administration times would be shorter than for the WMS-III. However, administration times reported in the *WMS-IV Administration and Scoring Manual* (Wechsler, 2009a, p. 14) indicated that for most participants, the total time for administering the WMS-IV Adult Battery was 75 to 77 minutes. The WMS-IV Older Adult Battery administration times were shorter, between 35 to 41 minutes for most participants. These administration times should be considered quite tentative since the data were derived from inexperienced examiners (James Holdnack, personal communication, January 1, 2008).

It is fair to assume that administration times would become faster with greater experience. Future studies will more precisely determine administration times among experienced examiners and for various clinical populations.

In the past, practitioners concerned with time efficiency used short forms of the WMS-III/WMS-R. For example, a three-subtest WMS-III short form consisted of Logical Memory, Verbal Paired Associates, and either Faces or Family Pictures and correlated at a .97 level with General Memory and Immediate Memory (Axelrod, Ryan, & Woodward, 2001). A two-subtest short form composed of Logical Memory and Verbal Paired Associates had a similar correlation of .96 with General Memory (and Immediate Memory). These two short forms accounted for 95% and 87% of the variance in General Memory and Immediate Memory, respectively (Axelrod & Woodward, 2000). Concerns with developing a formal short form resulted in the *WMS-III Abbreviated* (Wechsler, 2002b), which used four WMS-III subtests that could be used to calculate visual and auditory memory indexes. Clearly clinicians are concerned with time efficiency and will probably use various short forms of WMS-IV. There are some data that support shorter forms of the WMS-IV, including the use of

three subtests (and even two subtests) to accurately estimate immediate and delayed memory (J. B. Miller et al., 2012). These subtests include Logical Memory, Visual Reproduction, and Verbal Paired Associates.

The WMS-IV was designed to be a fairly comprehensive measure of memory as reflected in the five indexes (four for older adults). There are also numerous methods for displaying and analyzing various combinations of scores, including index scores, scaled scores, percentiles, confidence intervals, graphical displays, subtest variability within indexes, contrast scores between subtests, contrast scores between indexes, and comparisons between the WAIS-IV and WMS-IV. The numerous options for displaying and organizing scores are clearly an advantage in that they allow clinicians to extend the possible meanings of test scores. For example, clinicians can determine whether an examinee's long-term (delayed) memory is significantly higher/lower than his or her short-term (immediate) memory. The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009a) provides additional strategies for analysis, including forensic applications, considerations for older populations, demographically corrected norms, and information on whether changes in scores on repeat testing represent reliable change. Difficulties with the sheer number of options are the extensive time required for training and an increased possibility that clerical errors will occur during scoring (Hopwood & Richard, 2005; J. J. Ryan & Schnakenberg-Ott, 2003). In addition, the large number of comparisons increases the likelihood that seemingly meaningful differences will occur simply due to chance ("random" significance). Both scoring errors and random significance may result in incorrect interpretations, leading to poor patient care. Clinicians must take particular care to make sure their interpretations are accurate.

An important unanswered question with the WMS-IV is the extent to which it actually measures the various components of memory. Its divisions (and corresponding indexes) into visual, auditory, and working memories are well supported. However, the distinction between immediate and delayed memory is not as well supported. This difficulty was also noted on the WMS-III. As a result, clinicians who note differences between the Immediate and Delayed Indexes should seek further support based on other measures and relevant history. A related and important issue is that the various components of memory (and corresponding indexes) are likely to perform differently across various clinical populations and age groups. A final unanswered question in need of further exploration is the extent to which the WMS-IV indexes relate to aspects of everyday memory. This question is often crucial for clinicians; many referral questions ask such things as how much supervision the patient might need or whether the client can return to work. Early research in this area has been mixed, with the WMS-IV not consistently relating to functional status (Drozdzick & Cullum, 2011; Jung, 2014).

USE WITH DIVERSE GROUPS

Since the WMS-IV is a measure of cognitive abilities, many of the considerations that apply to the Wechsler intelligence scales also apply to the WMS-IV, including level of acculturation, language proficiency, ensuring rapport, encouraging optimal effort, paying particular attention to nontest information, and caution interpreting the meanings of verbal (auditory) versus nonverbal (visual) comparisons. However,

an important contrast between the Wechsler intelligence scales and the WMS-IV is that general ability measures, such as the Wechsler intelligence scales, are typically used to determine functional level compared to the general population. This is often the goal for psychoeducational assessments, assessing intellectual deficiency, and vocational assessments. In these situations, demographically adjusted norms are not recommended (see “Use with Diverse Groups” in Chapter 5), and the norms provided in the manuals should be adequate. In contrast, the WMS-IV is much more frequently used to determine neuropsychological diagnosis and level of impairment. In these cases, using demographically corrected norms is recommended (Heaton, Taylor, & Manly, 2003; E. Strauss, Sherman, & Spreen, 2006; Wechsler, 2002a). The main rationale here is that, rather than normative comparisons being made, comparisons are typically made between a client’s current status and a presumed premorbid level of functioning. Demographically corrected norms are more likely to give a more accurate estimate of premorbid level.

Research on the WMS-III found that the highest scores occurred among European Americans followed by Hispanics; the lowest scores were found among African Americans (Heaton et al., 2003; Heaton, Miller, Taylor, & Grant, 2004). Thus, using the norms provided in the WMS manual may result in overestimating the number of African Americans and other minorities who have “impairments.” Normative adjustment for age, education, gender, and ethnicity are thus available for the WMS-IV (Holdnack, Drozdick, Iverson, & Weiss, 2013; Pearson, 2009d). Clinicians should also take a careful history to make sure that ethnicity is indeed the actual variable that needs to be corrected. Other possibilities that might lower performance include quality of education, quality of the home environment, socioeconomic status, level/persistence of poverty, and health/nutritional status.

Some clients with physical, sensory, or language limitations might need special consideration with test administration and interpretation (see guidelines in Sattler, 2008, and E. Strauss et al., 2006). For example, it might be advisable to administer only the Auditory Memory and Symbol Span subtests to persons with physical difficulties (Pearson, 2009c). In contrast, it might be appropriate to give only the visual subtests and not the auditory (verbal) subtests to persons with language difficulties. If a client is not fluent in English, it might be advisable to administer the WMS-IV in a client’s native language. However, the advantages of greater comprehension should be balanced with the reduction in test validity resulting from nonstandardized administration. Any modification of test administration should be noted on the Record Form and in the psychological report.

INTERPRETATION PROCEDURE

The WMS-IV measures a wide range of different functions. As a result, interpretation can be complex. The strategies to be described focus on index scores and comparisons between various combinations of index scores. Since referral questions frequently ask how a patient’s memory compares with his or her overall ability, a section has been included on various relationships between scores on the WAIS-IV and WMS-IV. This interpretive approach is designed to focus on the most important dimensions of the

WMS-IV. More detailed information on a wider range of interpretive strategies can be found in the *WMS-IV Technical and Interpretive Manual* (Wechsler, 2009b) and *Essentials of WMS-IV Assessment* (Drozdzick, Holdnack, & Hilsabeck, 2011).

Psychological reports often include quite technical interpretative phrases. For example, a clinician might write something like “Ms. Example’s auditory immediate memory was statistically higher than her visual immediate memory.” This may be an accurate statement, but, at the same time, it is likely to be understood by a relatively narrow group of readers. Typically a much wider audience will read these reports. As a result, clinicians may wish to use an interpretive phrase, such as “Ms. Example’s ability to recall information that has been spoken to her was much better (top 50% of the population) than her ability to remember information she has seen (lowest 2% of the population).” This statement might be linked to actual test behavior, such as “She had a difficult time recalling details of designs she had been shown and then requested to draw.” Another option is to link test scores to examples of everyday behavior, such as “This suggests she would have problems remembering whom she had met previously or how she had gotten from one place to the next.”

Prior to administering the WMS-IV primary subtests, clinicians may choose to give the optional Brief Cognitive Screening Exam (see Table 6.1). This procedure presents clients with a series of fairly basic tasks, such as recalling the date/day/month, counting backward, drawing a clock, or naming objects they had previously been shown. A total score can be used to obtain a general sense of any major cognitive difficulties. Thus this exam serves a similar function as a mental status evaluation (see Chapter 3). Scores are converted into classification labels for Average (25%–100%), Low Average (10%–24%), Borderline (5%–9%), Low (2%–4%), and Very Low (<2%). As can be seen, scores are not so much geared toward high and superior levels of functioning but more toward distinguishing various levels of poor functioning. If a patient obtains a low or very low score, the clinician may even decide not to proceed with the more demanding tasks of the primary WMS-IV subtests.

Prior to interpreting the WMS-IV, clinicians should thoroughly understand these essential principles:

- The WMS-IV index and subtest scores are arranged in the same format as the WAIS-IV. The indexes have a mean of 100, with a standard deviation of 15. The range (floor and ceiling) extends from 4 standard deviations above the mean (160) to 4 standard deviations below the mean (40). Percentile ranks are calculated as part of the standard scoring procedure. Subtest scores have a mean of 10 and a standard deviation of 3 (range is 1–19).
- Whereas the index and subtest scores provide information on how the patient performs in relation to their age-related peers, “contrast scores” measure the differences between two scores and thus evaluate differences within the individual’s own functioning. One ability score is referred to as the “control,” since it becomes the basis of comparison; the second one is referred to as the “dependent” measure. For example, a clinician might wish to note whether patients’ memory for information they have seen (based on their Visual Memory Index) is significantly lower than for information they have heard (based on their Auditory Memory

Index). In this case, the Auditory Memory Index is the control measure; Visual Memory is being contrasted with it and thus is the dependent measure.

- Memory is a complex function that can be influenced by factors other than memory itself. These factors include poor hearing, language impairment, visual difficulties, poor attention, general intellectual impairment, and impaired executive functioning. Clinicians should always consider the possibility of comorbid conditions when trying to determine the reasons for difficulties in memory (e.g., traumatic brain injury being comorbid with a learning disability or the side effects of medication). It is incumbent on clinicians to identify whether low scores on the WMS-IV are due to specific problems with memory or are secondary to one or more other factors, such as those just listed.
- Patterns of WMS-IV scores cannot be used to diagnose specific clinical conditions. In other words, there is no WMS-IV score “fingerprint” that is specific to a given condition. However, when combined with other information, the WMS-IV can be a potentially crucial source of information to help with diagnosis.
- When the WMS-IV is used with diverse groups, clinicians should carefully consider clients’ acculturation and language facility. Proficiency with American English often allows examinees to more easily understand the directions and encode, consolidate, and retrieve the information. This is likely to be more important for auditory information than for more visually oriented tasks.
- Clinicians should be careful not to overinterpret WMS-IV scores. Sometimes overinterpretation can occur from relying on a single low subtest score. In fact, it is not unusual for average, healthy persons to have one low score (Brooks, Iverson, Holdnack, & Feldman, 2008). Another potential source of overinterpretation is to confuse statistical significance with clinical significance. In other words, just because a formal calculation has found that a low score is statistically significant does not mean that this indicates an “impairment,” “deficit,” or “pathology.” A closer inspection of cumulative percentages may reveal that many differences occur relatively frequently in the normal population. Thus, normal patterns of individual differences in memory abilities should not be confused with cognitive impairments. Finally, overinterpretation can potentially occur when a large number of scores are analyzed and some emerge as statistically significant, which might be mistaken as being clinically significant, though in reality they might merely be a random event (“random spurious significance”).
- Different clinicians may vary in their determinations of whether a score is “impaired.” As a general rule of thumb, WMS-IV index scores of 70 (2 standard deviations below the norm or within the second percentile) occurred in the most impaired clinical groups used in the validation studies (Alzheimer’s disease, mild/moderate intellectual disability). In contrast, borderline to low-average scores (70–85) occurred among clinical groups that were less impaired (patients with schizophrenia, patients who had had their temporal lobes removed, patients with moderate to severe traumatic brain injury). However, this impairment should also be considered within the context of a client’s overall abilities and occupation. For example, attorneys who rely heavily on auditory/verbal

skills may have considerable difficulty functioning in the profession if their auditory/verbal memory performance has been lowered into the average to low-average range.

- The main focus of interpretation should be on the index scores, which represent robust, psychometrically sound measures. In contrast, subtest scores are not as psychometrically sound. As a result, the next section offers minimal emphasis on subtest interpretation. Instead, subtests should be used to develop tentative hypotheses in need of further support (see Table 6.1). Subtests can also be used to make qualitative descriptions that can assist report readers to understand the types of behaviors on which the interpretations have been based (e.g., “For example, Mr. Example did poorly at recalling details of a brief short story that was read to him”).
- When minimal variability among the subtests occurs on an index, the index itself can be interpreted with a high level of confidence. In contrast, high subtest variability suggests that the unity of the index might be compromised due to potentially disparate abilities. This does not invalidate the index, but it does challenge clinicians to determine why there was less consistency in performance.

INTERPRETING PATTERNS OF INDEX SCORES

The purpose of interpreting patterns of index scores is to better understand a person’s memory-related strengths and weaknesses. Initially, clinicians might do this by noting the absolute values of the index scores. These values will provide comparisons with the standardization group. For example, a relatively low score on Visual Memory might indicate a relative weakness in this modality compared to the examinee’s age-related peers. Similarly, a low score on Auditory Memory might suggest difficulties with recalling verbally meaningful information. However, clinicians should also be aware that fluctuations can occur for a number of different reasons. It is up to each clinician to evaluate these various possibilities by carefully integrating additional relevant information. Therefore, the index “interpretations” listed in this section should be considered tentative.

Another strategy is to compare various combinations of index scores. Instead of making normative comparisons, this level of interpretation compares clients with their own relative strengths and weaknesses (so-called ipsative analysis). The comparisons included here are based on those distinctions that both are most clinically useful and have received empirical and theoretical support. Thus, a clinician may wish to know if visual or auditory modalities are relatively stronger or weaker. A second issue relates to whether a low score on visual working memory is really due to poor *working* memory or is rather the result of poor visual memory in general. A final assessment issue relates to differences between immediate (short-term) and long-term (delayed) memory. Thus, a delayed memory that is significantly lower than immediate memory suggests that there is likely a decay (forgetting) of memory over time. These three comparisons are described in the next section and can be formally determined by calculating and referring to the “Index-level contrast scaled scores” on the WMS-IV Record Form (and calculated by using conversion tables in Appendix G, Table G.12, of the *WMS-IV*

Administration and Scoring Manual). Knowledge related to each of these components of memory has relevance for diagnosis and treatment planning, as well as for understanding normal levels of strengths and weaknesses.

The next section briefly describes the index or contrast scores, then summarizes what a high or low score means along with some examples of how the abilities measured by the indexes might present in everyday life. Finally, consideration is given to understanding more in-depth aspects of the index, especially when there is wide variability or scores among the subtests.

Auditory Memory Index

The Auditory Memory Index (AMI) requires people to attend to information that has been presented to them orally. They then must comprehend the information and repeat it immediately after it has been presented. They must later recall the information again after a 20- to 30-minute delay. One subtest of the Auditory Memory Index (Logical Memory) requires examinees to repeat a brief story that has been read to them. The second subtest (Verbal Paired Associates) requests examinees to learn pairs of words that belong together (e.g., “dark ... *light*”; see Table 6.1).

High scores on the Auditory Verbal Index (AVI) indicate that the person has excellent abilities attending to and recalling information that he or she has heard. In contrast, low scores suggest the person will have difficulties attending to and recalling information that he or she has heard. Everyday examples might include recalling material presented in lectures, remembering oral directions, remembering conversations a few days later, recalling shopping items without the help of a list, or recalling phone numbers that the person has been told. Persons with low scores might benefit from writing down oral information. Assuming their visual memory is intact, they might also learn to translate the information into visual cues (visual reminders or “mind maps” of more complex information).

Although the WMS-IV norms provided in the administration/scoring and technical/interpretive manuals do not take into account sex differences, females typically perform better than males on auditory memory. Analysis of the 1997 WMS-III found that females had a mean of 10.58 versus a mean of 8.46 for males on the Verbal Paired Associates total recall scaled scores (Basso, Harrington, Matson, & Lowery, 2000), and more recent analysis on the WMS-IV has confirmed that women outperform men on the overall index (F. Pauls, Petermann, & Lepach, 2013). This effect is moderately strong and should therefore be considered when making interpretations.

One potentially useful behavioral observation is that excessive embellishment of stories on Logical Memory I and II may be a maneuver to compensate for or cover up difficulty remembering accurate information. Such embellishment may result in coherent elaboration or more illogical confabulations. A further behavioral observation is to note whether a client remembers primarily the global idea of the story as opposed to quite specific linear details. This result may suggest a global, holistic mode of processing as opposed to a more linear approach.

If all the scores for the subtests comprising Auditory Memory are relatively close together, interpretation of the Auditory Memory Index is fairly straightforward.

In some situations, there might be fluctuations among the different scores (see Table F.1 in the *WMS-V Administration and Scoring Manual* and Record Form for “Subtest-level differences within indexes”), which would then require careful consideration of why these scores were discrepant. This can best be done by considering the differences and similarities of the subtests. The Logical Memory tasks require examinees to recall information that has been read in a short story format. In contrast, Verbal Associates requires examinees to learn pairs of words (e.g., “high ... low”) over four consecutive learning trials. These two subtests are similar in that the information is not only for auditory recall but for auditory *verbal* information (rather than musical or other types of sounds). However, they are different in that Logical Memory requires examinees to learn more complex, verbally relevant, semantic information, whereas Verbal Paired Associates is for simple pairs of words and involves a prompt (one word is used as a prompt for the person to repeat the second/paired word). A discrepancy between these subtests might be explained by understanding the differences between them. For example, a relatively higher Logical Memory score suggests that the person is better able to attend to and consolidate more verbally meaningful information.

Another consideration when parsing discrepancies might be to note whether scores on the delayed portions of the subtests (Logical Memory II, Verbal Paired Associates II) are significantly higher or lower than the immediate versions (Logical Memory I, Verbal Paired Associates I). In other words, is the examinee’s short-term memory better or worse than his or her long-term memory? For example, if scores on the delayed versions were significantly lower, it suggests that examinees forget over time information that they have heard and initially learned. In order to help understand this distinction further, clinicians should note the Immediate and Delayed Memory indexes. In addition, they should obtain information from the client and informants to see if there are noteworthy examples of material that initially has been learned but seems to have been forgotten a short time later.

Visual Memory Index

The tasks on the Visual Memory Index (VMI) require examinees to recall designs from memory and either draw them or place them in the correct spatial location (see Table 6.1). As a result, it measures their memory for both visual details and where visual information should be located. Since examinees must respond to information both immediately after it has been presented and after a 20- to 30-minute delay, the Visual Memory Index is a measure of both short-term and long-term visual memory.

High scores on the Visual Memory Index suggest that examinees have good abilities in recalling the details and location of information they have seen. In contrast, low scores indicate problems with remembering the details and location of information they have seen. Everyday examples might include remembering whom they had seen earlier in the day, where something has been left in the house, how they had gotten from one place to the next, or finding where their car was parked in a parking lot. Patients with low scores might compensate by writing events that have occurred in a diary or writing down directions in a verbal form.

As with the Auditory Memory Index, it is useful to take into account gender differences in typical performance on the Visual Memory Index. Some research reveals

that, typically, males perform better than females on visual memory. Analysis on the WMS-IV revealed that men outperform women on the overall index (F. Pauls et al., 2013). This effect is moderately strong and should therefore be considered when making interpretations.

If all the subtest scores in the Visual Memory Index are similar, index interpretation can be fairly straightforward, as the abilities that have been measured are more likely to be unitary. Interpretation is more complicated in cases where there is wide variability among the subtest scores, suggesting that there may be quite specific visual memory difficulties (see Table F.1 in the *WMS-V Administration and Scoring Manual* and Record Form for “Subtest-level differences within indexes”). Visual Reproduction (I and II) requires examinees to look at a design and then draw it. Not only must they recall the design itself, but they must also go through the perceptual process of internally reconstructing it along with the external motor task of actually drawing the design. In contrast, Designs (I and II) requires examinees to look at the location of objects on a grid and later recall where the designs belonged on the grid. Thus the Designs subtest has more of a spatial component than Visual Reproduction. In contrast, Visual Reproduction has more of a psychomotor integration and reconstructive component. Discrepancies between scores on these subtests might be explained by understanding the differences in the tasks. For example, a significantly lower Visual Reproduction score compared with Designs scores might be due to examinees having difficulties with the task of having to draw the design. It should be noted that the Older Adult Battery does not include Designs I and II.

A further comparison might be made between the immediate and delayed portions of the subtests on Visual Memory. If scores for delayed visual memory tasks are significantly lower than those for immediate tasks, it suggests that the visual information that has been learned has been forgotten over time. Support for this, and other related inferences, should be obtained by noting performance on the Immediate and Delayed indexes, as well as obtaining information on the client’s everyday behavior. For example, do other people in the client’s life describe noteworthy instances in which the client seems to have rapidly forgotten information that has been seen (e.g., recalling who was at a meeting)?

Auditory Memory Index versus Visual Memory Contrast Scaled Score

One of the basic distinctions supported by WMS-IV factor analysis is between auditory and visual memory. As such, the difference between these modalities (and the indexes that measure them) can be used to hypothesize relative auditory versus visual strengths and weaknesses. It answers the question, “When auditory and visual memory abilities are compared, is one stronger or weaker, or are they similarly developed?” A significant difference can indicate either lifelong patterns related to differences in abilities or acquired deficits in these modalities. The WMS-IV converts differences between index scores into scaled scores with means of 10 and standard deviations of 3 (see *WMS-IV Administration and Scoring Manual*, Appendix G, Table G.12, and Record Form for “Index-level contrast scaled scores”), with the Auditory Memory Index as the control variable against which the Visual Memory Index as the dependent measure is compared. A scaled score of 7 (16th percentile) indicates that the examinee’s visual

memory is a weakness (of low-average magnitude) compared with his or her auditory memory. Lower scores exaggerate this difference and suggest visual memory impairment. In contrast, a score of 13 or greater (84th percentile) suggests that the examinee's visual memory is a relative strength compared with his or her auditory memory.

Some research has found hemispheric laterality differences in patients with visual versus auditory memory impairments. Specifically, laterality differences have been noted previously; patients with unilateral left-hemisphere damage have been found to perform more poorly with verbal-auditory information than with visual information (K. A. Hawkins, 1998; Pearson, 2009a). For example, these individuals would be expected to have particular difficulty when given verbal directions. In contrast, they might perform far better when shown a visual map of how to get from one place to the next. Patients with unilateral right-hemisphere damage would be expected to do more poorly on visual memory tasks. Thus, they would be expected to benefit more from auditory-verbal directions than from directions that were visually presented. However, visual memory performance was found to be the most sensitive to any type of brain damage, and patients with both unilateral right- and left-hemisphere damage performed poorly on visual memory types of tasks (Pearson, 2009a). If one modality was found to be relatively stronger than another, this stronger modality might be used to maximize learning. For example, if a person achieved a significantly lower score on auditory than visual memory tasks, he or she might use learning strategies that capitalized on visual modes (or vice versa).

Visual Working Memory

The Visual Working Memory Index (VWMI) assesses a person's ability to temporarily attend to, organize, and manipulate visuospatial information (see Table 6.1). Note that it is not included on the Older Adult Battery. Visual Working Memory is similar to the WAIS-IV Working Memory Index in that both indexes evaluate the degree to which a person can hold and manipulate information for a short period of time. However, the index on the WAIS-IV is specific to auditory-verbal material. It includes tests that require examinees to repeat and reorganize series of numbers and letters. In contrast, the WMS-IV has developed quite a different measure of working memory that is specific to visual information. The WMS-IV Visual Working Memory subtests require examinees to add/subtract visual information (Spatial Addition) and to arrange visual information into the correct sequence.

High scores on Visual Working Memory suggest that the person has excellent abilities holding and manipulating visual information. In contrast, low scores indicate the person has difficulties with these same visual tasks. Everyday examples might include being able to concentrate on a visual task without being distracted, staying focused on reorganizing furniture in a house, reorganizing the sequences of images on a computer screen, or tracking cards that have been seen in a card game.

As with the Visual Memory Index, males typically perform better than females on visual working memory. Again, some research on the WMS-IV revealed that men outperform women on the overall index (F. Pauls et al., 2013). This effect is moderately strong and should therefore be considered when making interpretations.

Interpretation of the Visual Working Memory Index is made easier when the subtest scores are all within the same range, which indicates that the ability is more unitary. In contrast, subtest scores that are quite variable suggest variations in more specific aspects of visual working memory (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for “Subtest-level differences within indexes”). As with previous indexes, it is incumbent on the clinician to parse these abilities in order to better understand the meaning of the index score. The Spatial Addition subtest requires examinees to look at two grids with different color circles. They must then add or subtract the location of the circles by following a set of rules. It is somewhat a spatial equivalent to the WAIS-IV Arithmetic subtest. Symbol Span shows examinees a series of abstract symbols on a page. They are then shown a different page with an array of symbols, including some from the previous page. They must identify which symbols had been shown to them previously and then indicate the order in which they were presented on the original page. The task is somewhat a visual analog to the WAIS-IV Digit Span subtest. Whereas Spatial Addition seems to be more of a visual “arithmetic” subtest (addition of the symbols is involved), Symbol Span involves more visual sequencing (items must be placed in the correct order). Clinicians should take these differences into account when understanding discrepant scores between the subtests. For example, a much lower Symbol Span subtest suggests that visual sequencing may be a particular difficulty for the examinee.

Visual Working Memory Index Versus Visual Memory Index Contrast Scaled Score

An important consideration in understanding an examinee’s performance on Visual Memory is whether it is due to poor memory itself or if poor visual working memory is impairing the memory process. In other words, is the problem a visual memory impairment beyond merely difficulties with working memory? Clinicians can determine the answer by checking to see if there is a significant difference between Visual Working Memory (the control variable) and Visual Memory (the dependent measure). The WMS-IV converts differences between index scores into scaled scores with means of 10 and standard deviations of 3 (see *WMS-IV Administration and Scoring Manual*, Appendix G, Table G.12, and Record Form for “Index-level contrast scaled scores”). A score of 7 (16th percentile) indicates that examinees’ visual memory is in the low-average range compared to their relatively higher visual *working* memory. If both are low compared to the normative sample, this finding suggests that their visual memory is impaired even beyond what poor visual working memory may cause. For example, patients who have had their right temporal lobes removed were found to have contrast scores of 7.7 (Pearson, 2009a), which reflects even poorer visual memories than visual working memories. This somewhat low score is what would be expected, given that the right temporal lobe processes information related to visual memory. In contrast, a scaled score of 13 (84th percentile) suggests that clients have higher visual memory relative to their visual *working* memory. It can be concluded from high scores like this that visual working memory may be limiting ability in visual memory. In such cases, more faith can be placed in the interpretation that it was visual *working* memory itself that caused the low performance.

Sometimes scores on both the Visual Working Memory and Visual Memory indexes are low. In these cases, the contrast score is likely to be average. This situation is likely caused by poor visual memory processes in general. It may also reflect interference on the tasks due to impaired visual perception.

Immediate Memory Index

Short-term (immediate) and long-term (delayed) memory are two of the crucial distinctions related to understanding memory. The WMS-IV Immediate Memory Index (IMI) assesses how well examinees can recall both verbal and visual information immediately after the information has been presented. It includes tasks that require examinees to recall a story that has been read to them, learn words that are paired, draw designs from memory, and recall the correct location where designs should be placed on a grid (see Table 6.1).

High scores suggest that clients have good short-term memory for recalling information that they have heard (auditory-verbal) and information that they have seen (visual). In contrast, low scores indicate that clients have difficulty with these abilities. Everyday examples might include being able to recall a license plate they have seen, a phone number they have been told to remember, or names of people at a party. However, these examples all relate to information that clients can recall on a short-term (immediate) basis. It does not necessarily imply that they will be able to recall it over a longer duration.

Interpretation of the Immediate Memory Index is relatively clear when all the subtests scores are fairly even. This means that the ability being measured is unitary. In contrast, variation among the subtests means that the overall score might have occurred due to more specific factors (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for “Subtest-level differences within indexes”). The most obvious factor that can affect the overall Immediate Memory Index score is differences in auditory as opposed to visual abilities. As such, clinicians should note scores on the Auditory and Visual Memory Indexes, as well as the Auditory Memory Index versus Visual Memory Contrast Scaled Score. For example, if the Immediate Memory Index score was low but the Visual score was much lower than the score for the Auditory Index, it suggests that the relatively poor performance on visual material was mainly responsible for the poor performance on the Immediate Memory Index.

Delayed Memory Index

In addition to measuring short-term (immediate) memory, the WMS-IV also measures the extent to which examinees retain information for a slightly longer period of time. This is measured by requesting examinees to recall details of the information that has been presented to them in each of the primary subtests following a 20- to 30-minute delay. Clients must first attend to the information and then encode, consolidate, retrieve, and provide the correct answer after the delay.

Persons with high scores on the Delayed Memory Index (DMI) can be expected to be good at retaining and retrieving information they have learned. In contrast, persons with low scores can be expected to have difficulty retaining and retrieving

information. Everyday examples might include long-term recall of instructions, times of meetings, where things should be placed in cupboards, and repeating jokes or stories they have heard.

The Delayed Memory Index is comprised of many memory components, as it requires that clients first accurately encode and consolidate short-term (“immediate”) information and then recall it at a later time. It involves both visual and auditory information. As a result, it can be conceptualized as a measure of global memory (similar to the General Memory Index on the 1997 WMS-III; James Holdnack, personal communication, January 6, 2009).

As with the previous indexes, subtest scores on the Delayed Memory Index that are similar mean that the ability is likely a unitary construct. As a result, the index can be interpreted with confidence. In contrast, a high level of variability challenges the clinician to determine if there are more specific abilities that resulted in differential performance (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for “Subtest-level differences within indexes”). Doing this can be particularly challenging since the Delayed Memory score represents the end product of a wide array of processes (attending to the test material, good visual/auditory perception, encoding, consolidation, retrieval, expressing the response). As a result, a wide variety of issues can disrupt performance on Delayed Memory. One potentially important distinction is between visual and auditory-verbal modalities. Clinicians might inspect possible differences in subtest scores to determine if the visual subtest scores are higher/lower than the auditory-verbal scores. They might also check the Visual and Auditory Index scores and note especially the Auditory Memory Index versus Visual Memory Contrast Scaled Score. If, for example, the auditory scores were significantly lower than the visual scores, it suggests that the client’s auditory memory might be driving the low score on the Delayed Memory Index.

One possible explanation for a low Delayed Memory score might be that people had a difficult time retrieving the information, even if they had learned it successfully. Thus they might not have been able to “recall” the correct answers, but, if given a chance, they might be able to “recognize” the correct answers. This finding can be parsed by having administered the recognition procedures (see WMS-IV Record Form “Process Score Conversion” section). They might have scored poorly in the standard scores on the primary subtests, but their recognition scores might have been quite good. Everyday demonstration of this discrepancy might be persons who have extensive tip-of-the-tongue struggles, say “I know I know the answer, but I just can’t remember it,” or can get the correct answer with minor prompting.

Immediate Memory Index Versus Delayed Memory Index Contrast Scaled Score

The distinction between short-term (immediate) and long-term (delayed) memory is often important for practicing clinicians. Thus, a referral question might be: “Does this patient have impairments in forgetting material she has previously learned?” In contrast, the memory of some examinees actually might improve over time because they need the extra time to consolidate the information. A question relating to this issue might be: “Does the examinee have an improvement in memory over time?” If delayed

memory is considerably lower than immediate memory (see *WMS-IV Administration and Scoring Manual*, Appendix G, Table G.12, and Record Form for “Index-level contrast scaled scores”), it suggests that the person can learn material initially but that the information decays over a period of time. It should be stressed that performance on immediate memory becomes the benchmark for how much information has been lost. In other words, unless a person has learned something initially, there is nothing to lose. A variation might be that a person has acquired information but then may not be able to recall it due to poor retrieval. However, recognizing information is generally a much easier task; the person might be able to recognize information accurately even though he or she may not be able to recall/retrieve that information. A number of procedures are available on the WMS-IV to contrast a person’s recall with recognition (see Record Form for relevant “Process Score Conversions” and “Subtest-Level Contrast Scaled Scores” scores).

One issue is that factor analysis of the immediate/delayed distinction on the WMS-IV is not as strong as would be optimal (Pearson, 2009d). This issue is consistent with the finding that there was a quite high (.87) correlation between the Immediate Memory and Delayed Memory Indexes, as was also the case for the WMS-III (K. A. Hawkins, 1998; Millis et al., 1999; Weiss & Price, 2002). Clinically this means that most of the time, the index scores will not reveal a significant difference between the two abilities. One potential reason for this is that the relatively short time difference between immediate and delayed recall (20–30 minutes) may not be enough time to distinguish these actual abilities. Despite these findings, it was decided to include Immediate and Delayed Memory indexes on the WMS-IV, since they can still provide potentially useful clinical information (Millis et al., 1999). In other words, there may be some populations (e.g., those with Korsakoff’s disease, older populations) who can repeat information they have just seen or heard but forget it a short time later. This situation can be suspected in cases where informants state that clients seem to understand and can repeat information but cannot say what they saw or heard the next day.

COMPARING SCORES ON THE WAIS-IV AND THE WMS-IV

One of the most important referral questions is whether a client’s memory is lower than would be expected, given his or her other, more general abilities. The question may be phrased in this way: “Is this client’s memory consistent with his general level of cognitive functioning?” A comparison between performance on the WMS-IV and performance on the WAIS-IV allows a clinician to answer this question. It places memory performance within a larger context. Thus, general ability (WAIS-IV) provides a baseline or comparison point for evaluating the extent that memory (on the WMS-IV) has potentially declined. This assumes, of course, that the more general abilities measured on the WAIS-IV are relatively stable. In contrast, memory is usually considered to be more sensitive to decline, a finding consistent with the fact that memory is often a major presenting concern for patients. For example, memory is usually the main complaint reported by patients with traumatic head injury or Alzheimer’s disease. In contrast, their other more general abilities tend to be more stable.

Previous versions of the WMS used a total or general score that could be compared with a person's WAIS Full Scale IQ. The original WMS allowed practitioners to calculate a "Memory Quotient," and the 1997 WMS-III had a General Memory Index. Differences between the general ability and the memory scores were fairly easy to explain to referral sources or family members. Clinicians could use a phrase such as "Joe's overall mental abilities were in the average range (50th percentile), but, in contrast, his memory was much lower since he was in the bottom 5% of the population." Instead, the WMS-IV makes comparisons between each of its index scores and the WAIS-IV General Ability Index (GAI). The General Ability Index was used since it is comprised of verbal (Verbal Comprehension Index) and performance/nonverbal (Perceptual Reasoning Index) abilities, which tend to be fairly resistant to the impact of most clinical disorders and situational factors (like lack of sleep). As such, they are quite stable. In contrast, speed (Processing Speed Index) and attention/manipulation (Working Memory Index) are quite sensitive to a variety of clinical conditions and situational factors. Many of the conditions that would lower memory would also be likely to lower speed and attention (Processing Speed and Working Memory indexes). The General Ability Index is generally a more stable benchmark for comparison than the Full Scale IQ, which includes all four WAIS-IV indexes (including measures of speed and attention). In other words, the General Ability Index–WMS-IV index comparisons are likely to be more sensitive to difficulties with memory when compared to using Full Scale IQ–WMS-IV index comparisons.

The Record Form allows for completion of a wide number of ability-memory comparisons (see the Record Form "Ability-Memory Analysis" section that uses Tables B.1–B.16 on pp. 200–218 of the *WMS-IV Technical and Interpretive Manual*). Most clinicians focus primarily on comparisons between the General Ability Index and the WMS-IV index scores. For that reason, we describe only those scores in this chapter. Some examiners may want to make more detailed comparisons between additional combinations of the WMS-IV index and WAIS-IV index scores. All differences are converted to contrast scaled scores with a mean of 10 and a standard deviation of 3, with the ability (WAIS-IV) score representing the control variable and the memory index (WMS-IV) score representing the dependent measure. A score of 7 (16th percentile) indicates that the memory index is unexpectedly low (a relative weakness) compared with the General Ability Index. Lower scores exaggerate this difference and suggest the possibility of memory impairment specific to the index. In contrast, a score of 13 or greater (84th percentile) suggests that the examinee's memory index is unexpectedly high (a relative strength) compared with his or her General Ability Index. One caution is that, with so many potential comparisons, the possibility of random spurious significant differences increases. In other words, some of the "significant" differences may not actually be clinically accurate descriptions of the client. As a result, clinicians should be careful not to overinterpret the difference scores.

The next descriptions are quite brief. Clinicians who want more detailed interpretations can read information under each of the WMS-IV indexes; that material includes a description of the index, a listing of the types of tasks involved, a brief interpretation of the meaning of high/low scores, and everyday examples.

General Ability Index Versus Auditory Memory Index

A low score (below 7) indicates that the information that clients have recalled based on having heard it is a weakness compared with their overall ability. High scores (above 13) indicate that their memory for information they have heard is a relative strength.

General Ability Index Versus Visual Memory Index

A low score (below 7) indicates that the information that clients have recalled based on having seen it is a weakness compared with their overall ability. This comparative index has been found to be one of the most sensitive measures of impairment (Pearson, 2009d). High scores (above 13) indicate that clients' memory for information they have seen is a relative strength.

General Ability Index Versus Visual Working Memory Index

A low score (below 7) indicates that clients' ability to concentrate on, hold, organize, and manipulate complex visual information is a relative weakness compared with their overall ability. They are likely to have difficulty working with both where the information was located ("visual space") and the details of what was seen ("visual details"). High scores (above 13) indicate that holding and manipulating visual information (both spatially and for details) is a relative strength. It may also be important to compare the WAIS-IV Working Memory Index, which is a measure of auditory-verbal working memory, to the WMS-IV Visual Working Memory Index.

General Ability Index Versus Immediate Memory Index

A low score (below 7) indicates that clients' short-term ("immediate") memory for information they have seen or heard is a relative weakness compared with their overall ability. High scores (above 13) indicate that clients' short-term ("immediate") memory is a relative strength.

General Ability Index Versus Delayed Memory Index

A low score (below 7) indicates that clients' long-term ("delayed") memory for information they have seen or heard is a relative weakness compared with their overall ability. This measure is one of the more clinically sensitive measures (Pearson, 2009a). High scores (above 13) indicate that clients' long-term ("delayed") memory for information they have seen or heard is a relative strength. "Long"-term assessment on this index was based on a 20- to 30-minute delay. Since the Delayed Memory Index is comprised of many memory components, it can be conceptualized as a measure of global memory (similar to the General Memory Index on the 1997 WMS-III). Given the sensitivity of the Delayed Memory Index, combined with the fact that it is a global measure of memory, this index should be one of the most important comparisons.

ADDITIONAL CONSIDERATIONS: MALINGERING AND EVALUATING CHANGE

Secondary gain is frequently an issue for assessments related to personal injury litigation, workers' compensation, long-term disability, or defendants in criminal injury proceedings. Due to the potential for gain, malingering is a distinct possibility. Surveys of neuropsychologists found that estimates for feigning deficits were as high as 30% among personal injury and workers' compensation cases (Mittenberg, Patton, Canyock, & Condit, 2002) and up to 40% for litigants involved with traumatic brain injury (Larrabee, 2005). Memory problems are particularly likely to be exaggerated since they are often the most frequently reported problems among these populations. As a result, clinicians need to be particularly careful to evaluate the validity of a client's complaints. Terms that are similar to malingering but somewhat more neutral include *suboptimal engagement*, *inconsistent effort*, or *feigning*.

A number of specialty instruments are available to detect suboptimal cognitive effort and are recommended to help make a more definitive assessment (see K. A. Boone, 2007; Larrabee, 2005; E. Strauss et al., 2006). Best practice requires multiple measures to be used. Possible strategies to detect malingering on the WMS-IV might be to focus on the Logical Memory Delayed Recognition task that requests clients to state whether (yes or no) an item was included in one of the previously read stories. Because random guessing would produce a score of 50%, scores of less than this suggest that the client is purposely giving incorrect responses (see Killgore & Dellapietra, 2000). Malingering may also be suggested if recognition is not superior to poor recall, as recognition tasks are easier than free recall tasks. A final quite general indicator is dramatic differences between a person's day-to-day functioning (based on corroborating sources) and performance on WMS-IV measures. The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009a) provides additional strategies to detect malingering, including analyses of guessing for Logical Memory Recognition, Verbal Paired Associates, Designs Spatial, and Designs Content. If an examinee performs lower than guessing on the listed subtests, malingering is a possibility.

Sometimes WMS-IV scores are used to document deterioration or to monitor improvement. It is tempting to peruse pretest and posttest scores and quickly infer that some sort of actual change has occurred in the patient's level of functioning. For example, a client might have had a WMS-IV Delayed Memory Index score of 80 directly after a head injury and, three months later, achieved a score of 85. It might be inferred that the patient's memory has improved. However, this finding does not take into consideration factors such as practice effects, regression to the mean, or the relative reliability of the measure. The improvement between the pretest of 80 and the posttest of 85 might simply be the result of the patient's practicing the tasks three months previously, or the difference might simply be measurement error (reflected in its test-retest reliability). The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009a) provides strategies for calculating whether actual change has occurred ("reliable change index"). These calculations account for the imperfect reliability of the instrument, which helps determine whether significant change in

scores has been demonstrated or not, although this determination may not necessarily relate to personal or social significance of the change (see Beutler & Moleiro, 2001). Determining the personal and clinical meaning of changed scores requires clinicians to integrate information from a wider variety of sources to support any inferences related to actual change in the client's functioning.

RECOMMENDED READING

- Holdnack, J. A., Drozdick, L., Iverson, G. L., & Weiss, L. G. (2013). *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation*. San Diego, CA: Academic Press.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). Memory. In O. Spreen & E. Strauss (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 679–686). New York, NY: Oxford University Press.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY[†]

The Minnesota Multiphasic Personality Inventory (MMPI)¹ is a standardized questionnaire that elicits a wide range of self-descriptions scored to give a quantitative measurement of an individual's level of emotional adjustment and attitude toward test taking. Since its original development by Hathaway and McKinley in 1940, the MMPI has become the most widely used clinical personality inventory, with more than 10,000 published research references (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Boccaccini & Brodsky, 1999; Camara, Nathan, & Puente, 2000; C. Piotrowski, 1999). Thus, in addition to its clinical usefulness, the MMPI has stimulated a vast amount of literature. Currently, two different versions of the test are available and in use, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the MMPI-A Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011), each which has strengths and limitations. Both are presented in this chapter.

The 1943 MMPI test format consisted of 504 affirmative statements that could be answered “True” or “False.” The number of items was later increased to 566 through the inclusion of repeat items and Scales 5 (Masculinity-Femininity) and 0 (Social Introversion). The 1989 restandardization retained the same basic format but altered, deleted, and/or added a number of items, which resulted in a total of 567 items. The different categories of responses can be either hand- or computer-scored and summarized on a profile sheet. An individual's score as represented on the profile form can then be compared with the scores derived from different normative samples.

The original MMPI had 13 standard scales, of which 3 related to validity and 10 related to clinical or personality indices. The more recent MMPI-2 and MMPI-A (the adolescent version) have maintained the original 10 clinical/personality scales as well as the original 3 validity scales, but the total number of validity scales has been increased (see Table 7.1). The MMPI-2-RF has developed new, though many theoretically analogous, clinical and validity scales, in addition to adding 3 higher order scales (see Table 7.1). The MMPI-2 clinical and personality scales are known both by their scale numbers and by scale abbreviations. On both the MMPI-2 and MMPI-2-RF, additional options are available to refine the meaning of the clinical scales as well as provide

[†]For Client Feedback Statements denoted by a dagger: Copyright 2011 from *Therapeutic Feedback with the MMPI-2: A Positive Psychology Approach*, by R. W. Levak, L. Siegel, & D. S. Nichols, 2011. Reproduced by permission of Taylor & Francis, Inc., permission conveyed through Copyright Clearance Center, Inc.

¹MMPI-2™ (Minnesota Multiphasic Personality Inventory–2)™ Test Booklet. Copyright © 1942, 1943 (renewed 1970), 1989 by the Regents of the University of Minnesota. All rights reserved. Used by permission of the University of Minnesota Press. “MMPI-2” and “Minnesota Multiphasic Personality-2” are trademarks owned by the Regents of the University of Minnesota.

Table 7.1 Validity, Higher-Order, and Basic (Clinical and Restructured Clinical) MMPI-2 and MMPI-2-RF Scales

MMPI-2	Abbreviation (scale no.)	No. of items	MMPI-2-RF	Abbreviation	No. of items
Validity scales					
Cannot say	?		Cannot say	CNS	
Variable response inconsistency	VRIN	98	Variable response inconsistency	VRIN-r	106
True response inconsistency	TRIN	40	True response inconsistency	TRIN-r	52
Infrequency	F	60	Infrequent responses	F-r	32
Back F	Fb	40			
Infrequency- psychopathology	Fp	27	Infrequent psychopathology responses	Fp-r	21
Fake Bad Scale	FBS	43	Symptom validity	FBS-r	30
			Response bias scale	RBS	28
Lie	L	15	Uncommon virtues	L-r	14
Correction	K	30	Adjustment validity	K-r	14
Superlative self presentation	S	50			
Higher-order scales					
			Emotional/internalizing dysfunction	EID	41
			Thought dysfunction	THD	26
			Behavioral/externalizing dysfunction	BXD	23
Basic (clinical) scales					
			Demoralization	RCd	24
Hypochondriasis	Hs (1)	32	Somatic complaints	RC1	27
Depression	D (2)	57	Low positive emotions	RC2	17
Hysteria	Hy (3)	60	Cynicism	RC3	15
Psychopathic deviate	Pd (4)	50	Antisocial behavior	RC4	22
Masculinity-femininity	Mf (5)	56			
Paranoia	Pa (6)	40	Ideas of persecution	RC6	17
Psychasthenia	Pt (7)	48	Dysfunctional negative emotion	RC7	24
Schizophrenia	Sc (8)	78	Aberrant experiences	RC8	18
Hypomania	Ma (9)	46	Hypomanic activation	RC9	28
Social introversion	Si (0)	69			

additional information. These options include scales based on item content (content or problem scales), subscales for the clinical and personality scales based on clusters of content-related items (Harris-Lingoes subscales), assessment of items and item clusters that relate to relevant dimensions (critical items), personality psychopathology scales (PSY-5 scales), and empirically derived new scales (supplementary scales). New scales are continually being researched and reported in the literature. The result of these developments is an extremely diverse and potentially useful test that can be interpreted, refined, and expanded from a variety of different perspectives.

The contents for the majority of MMPI questions are relatively obvious and deal largely with psychiatric, psychological, neurological, or physical symptoms. However, some of the questions are psychologically obscure because the underlying psychological process they are assessing is not intuitively obvious. For example, item 68, "I sometimes tease animals," is empirically answered "False" more frequently by depressed subjects than normals. Thus, it was included under Scale 2 (Depression) even though it does not, on the surface, appear to directly assess an individual's degree of depression. For the most part, however, the statements are more direct and self-evident, such as item 56, "I wish I could be as happy as others seem to be," or 146, "I cry easily," both of which also reflect an examinee's level of depression. The overall item content is extremely varied and relates to areas such as general health, occupational interests, preoccupations, morale, phobias, and educational problems.

After a test profile has been developed, the scores are frequently arranged or coded in a way that summarizes and highlights significant peaks and valleys. However, to interpret the test accurately, both the overall configuration of the different scales and the relevant demographic characteristics of the client must be considered. In many instances, the same scaled score on one test profile can mean something quite different on another person's profile when the elevations or lowerings of other scales are also considered. For example, an elevated Scale 6/RC6 (Paranoia/Ideas of persecution) may indicate an individual who feels victimized, criticized, and judged and is highly sensitive to the actions of others. However, if this elevation is accompanied by a high 4/RC4 (Psychopathic deviate/Antisocial behavior), a likelihood of acting out against others is indicated. This acting out can seem sudden and inappropriate, often prompted by a stimulus that an outside observer would consider ambiguous or even neutral. Thus, it is important for the clinician to avoid the use of purely quantitative or mechanical formulas for interpreting the profile and instead examine the scores in the overall context of the other scale elevations and lowerings. A particular scale should be examined not only in the context of the overall test configuration, but also using additional sources, such as demographic characteristics (e.g., age, education, socioeconomic status, ethnicity), behavioral observations, other psychometric devices, and relevant history, which can often increase the accuracy, richness, and sensitivity of personality descriptions.

A further important, general interpretive consideration is that the scales represent measures of personality traits rather than simply diagnostic categories. Although the scales were originally designed to differentiate normal from abnormal behavior, it is generally regarded as far more useful to consider that the scales indicate clusters of personality variables. For example, Scale 2 (Depression) may suggest characteristics such as mental apathy, self-deprecation, and a tendency to worry over even relatively

small matters. This approach characterizes the extensive research performed on the meanings of the two highest scales (2-point code types), which are summarized later in this chapter. Rather than merely labeling a person, this descriptive approach creates a richer, more in-depth, and wider assessment of the individual who is being tested.

HISTORY AND DEVELOPMENT

The original development of the MMPI was begun in 1939 at the University of Minnesota by Starke R. Hathaway and J. Charnley McKinley. They wanted an instrument that could serve as an aid in assessing adult patients during routine psychiatric case workups and that could accurately determine the severity of their disturbances. Furthermore, Hathaway and McKinley were interested in developing an objective estimate of the change produced by psychotherapy or other variables in a patient's life.

The most important approach taken during construction of the MMPI was empirical criterion keying. This refers to the development, selection, and scoring of items within the scales based on some external criterion of reference. Thus, if a clinical population was given a series of questions to answer, the individuals developing the test would select questions for inclusion or exclusion based on whether this clinical population answered differently from a comparison group. Even though a theoretical approach might be used initially to develop test questions, the final inclusion of questions would not be based on this theoretical criterion. Instead, test questions would be selected based on whether they were answered in a direction different from a contrasted group. For example, a test constructor may believe that an item such as "Sometimes I find it almost impossible to get out of bed in the morning" is a theoretically sound statement to use in assessing depression. However, if a sample population of depressed patients did not respond to that question differently from a normative group, the item would not be included. Thus, if persons with hysterical traits generally answer "True" to the statement "I have stomach pains," especially more so than a group without hysterical traits, whether they actually do have stomach pains is less important, from an empirical point of view, than the fact that they say they do. In other words, the final criterion for inclusion of items in an inventory is based on whether these items are responded to in a significantly different manner by a specified population sample.

Using this method, Hathaway and McKinley began with an original item pool of more than 1,000 statements derived from a variety of different sources, including previously developed scales of personal and social attitudes, clinical reports, case histories, psychiatric interviewing manuals, and personal clinical experience. Of the original 1,000 statements, many were eliminated or modified. The result was 504 statements that were considered to be clear, readable, not duplicated, and balanced between positive and negative wording. The statements themselves were extremely varied and were purposely designed to tap as wide a number of areas in an individual's life as possible. The next step was to select different groups of nonclinical ("normal") and psychiatric patients to whom the 504 questions could be administered. The normals were primarily friends and relatives of patients at the University of Minnesota hospitals who were willing to complete the inventory. They consisted of 226 males and 315 females, who were screened with several background questions about age, education, marital status,

occupation, residence, and current medical status. Individuals who were under the care of a physician at the time of the screening were excluded from the study. This group was further augmented by the inclusion of other normal subjects, such as recent high school graduates, Work Progress Administration (WPA) workers, and medical patients at the University of Minnesota hospitals. This composite sample of 724 individuals was closely representative in terms of age, sex, and marital status of a typical group of individuals from the Minnesota population, as reflected in the 1930 census. The clinical group included patients who represented the major psychiatric categories being treated at the University of Minnesota hospitals. These patients were divided into clear subgroups of approximately 50 in each category of diagnosis. If a patient's diagnosis was at all in question, or if a person had multiple diagnoses, he or she was excluded from the study. The resulting subgroups were hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, psychasthenia, schizophrenia, and hypomania.

After the normals and psychiatric patients had been administered the 504-item scale, Hathaway and McKinley could then compare their responses. Each item that correctly differentiated between these two groups was included in the resulting clinical scale. For example, an item such as "I feel vague aches and pains in my stomach" might be answered "True" by 20% of a sample of those with hypochondriasis and by only 2% of the normals. It could thus be included in the clinical scale for hypochondriasis. The comparisons, then, were between each clinical group and the group of normals rather than among the different clinical groups themselves. This selection procedure was used to develop tentative clinical scales.

Still another step was included in the scale constructions. The fact that an item was endorsed differently by the group of 724 Minnesota normals than by the patients from various clinical populations did not necessarily indicate that it could be used successfully for clinical screening purposes. Thus, an attempt was made to cross-validate the scales by selecting a new group of normals and comparing their responses with a different group of clinical patients. The items that still provided significant differences between these groups were selected for the final version of the scales. It was reasoned, then, that these items and the scales composed of these items would be valid for differential diagnosis in actual clinical settings.

Whereas this procedure describes how the original clinical scales were developed, two additional scales that used slightly different approaches were also included. Scale 5 (Masculinity-Femininity) was originally intended to differentiate male homosexuals from males with a more exclusively heterosexual orientation. However, few items were found that could perform this function effectively. The scale was then expanded to focus on items that were characteristically endorsed in a certain direction by the majority of males and in the opposite direction by the majority of females. This was accomplished in part by the inclusion of items from the Terman and Miles I Scale (1936). The second additional scale, Social Introversion (*Si*), was developed by Drake in 1946. It was developed initially by using empirical criterion keying in an attempt to differentiate female college students who participated extensively in social and extracurricular activities from those who rarely participated. It was later generalized to reflect the relative degree of introversion for both males and females.

It soon became apparent to the test constructors that persons could alter the impression they made on the test because of various test-taking attitudes. Hathaway and

McKinley thus began to develop several scales that could detect the types and magnitude of the different test-taking attitudes most likely to invalidate the other clinical scales. Four scales were developed: the Cannot say (?) scale, Lie (*L*), Infrequency (*F*), and Correction (*K*) scales. The Cannot say (?) scale is simply the total number of unanswered questions. If a high number of these are present, it would obviously serve to reduce the validity of the overall profile. High scores on the Lie scale indicate a naive and unsophisticated effort on the part of the examinee to create an overly favorable impression. The items selected for this scale were those that indicated a reluctance to admit to even minor personal shortcomings. The Infrequency (*F*) scale is composed of those items endorsed by fewer than 10% of normals. A high number of scorable items on the *F* scale, then, reflects that the examinee is endorsing a high number of unusually deviant responses.

Correction (*K*), which reflects an examinee's degree of psychological defensiveness, is perhaps the most sophisticated of the validity scales. The items for this scale were selected by comparing the responses of known psychiatric patients who still produced normal MMPIs (clinically defensive) with "true" normals who also produced normal MMPIs. Those items that differentiated between these two groups were used for the *K* scale. Somewhat later, the relative number of items endorsed on the *K* scale was used as a "correction" factor. The reasoning behind this was that if some of the scales were lowered because of a defensive test-taking attitude, a measure of the degree of defensiveness could be added into the scale to compensate for this. The result would theoretically be a more accurate appraisal of the person's clinical behavior. The scales that are not given a *K* correction are those whose raw scores still produced an accurate description of the person's actual behavior. However, there have been some questions regarding the effectiveness of the *K* correction in some settings. As a result, clinicians have the choice of whether they wish to use MMPI-2 profile sheets with or without the *K* correction, and the MMPI-A has omitted the use of the *K* correction altogether.

Since the publication of the original MMPI, special scales and numerous adjunctive approaches to interpretation have been developed. A primary strategy has been content interpretation. The most frequently used have been the Harris and Lingoes subscales, Wiggins Content Scales, and several different listings of critical items. These scales can potentially provide important qualitative information regarding an examinee. In addition, many supplementary scales have been developed, such as the Anxiety Scale, the MacAndrew Scale to assess the potential for substance abuse, and the Ego Strength Scale to estimate the extent to which a person will benefit from insight-oriented therapy. Each of these approaches can be used as an adjunct to interpreting the traditional clinical scales and/or experimental scales for assessing or researching specific populations (Butcher, 2006, 2011; Butcher, Graham, Williams, & Ben-Porath, 1990; J. R. Graham, 2011; C. L. Williams, Butcher, Ben-Porath, & Graham, 1992).

In addition to innovations in scales and interpretations, the MMPI has been used in a wide number of settings for extremely diverse purposes. Most studies have focused on the identification of medical and psychiatric disorders, as well as on uses in forensic contexts (Deardorff, 2000; Greene, 2000; Pope, Butcher, & Seelen, 2000) and on expanding or further understanding the psychometric properties of the MMPI. Other frequent topics include alcoholism, aging, locus of control, computer-based interpretation, chronic pain, and the assessment of different occupational groups. The MMPI

has been translated into a number of different languages and has been used in a wide range of different cross-cultural contexts (see Butcher, 1996, 2004; Cheung & Ho, 1997; Greene, 1991; G. C. N. Hall, Bansal, & Lopez, 1999; Handel & Ben-Porath, 2000).

Criticisms of the original MMPI primarily centered on its growing obsolescence, difficulties with the original scale construction, inadequacy of its standardization sample, and difficulties with many of the items (Helmes & Reddon, 1993). Problems with the items included sexist wording, possible racial bias, archaic phrases, and objectionable content. In addition, the original norms had poor representation of minorities and were inappropriate in making comparisons with current test takers. Further problems have related to inconsistent meanings associated with *T* score transformations.

These criticisms led to an extensive restandardization of the MMPI, which began in 1982. Despite the need to make major changes, the restandardization committee wanted to keep the basic format and intent of the MMPI as intact as possible so that the extensive research base collected over the 50 prior years would be applicable to the restandardized version. As a result, these six goals were established:

1. Deletion of obsolete or objectionable items
2. Continuation of the original validity and clinical scales
3. Development of a wide, representative normative sample
4. Norms that would most accurately reflect clinical problems and would result in a uniform percentile classification
5. Collection of new clinical data that could be used in evaluating the items and scales
6. Development of new scales

The restandardization used a special research form consisting of the original 550 items (of which 82 were modified) and additional 154 provisional items used for the development of new scales. Even though 82 of the original items were reworded, their psychometric properties were apparently not altered (Ben-Porath & Butcher, 1989). The resulting 704-item form (Form AX) was administered to 1,138 males and 1,462 females from seven different states, several military bases, and a Native American reservation. The subjects were between the ages of 18 and 90 and were contacted by requests through direct mail, advertisements in the media, and special appeals. The resulting restandardization sample was highly similar to the 1980 U.S. census in almost all areas, with the exception that the sample was somewhat better educated than the overall population.

The MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) differs from the older test in a number of ways. The *T* scores that subjects obtain are generally not as deviant as those from the earlier version. In addition, the *T* scores were designed to produce the same range and distribution throughout the traditional clinical scales (except for Scales 5 and 0). The practical result is that *T* scores of 65 or greater are considered to be in the clinical range (versus a cutoff score of 70 for the MMPI). Also, the percentile distributions are uniform throughout the different clinical scales (whereas they were unequal for the MMPI). The test booklet itself contains 567 items, but the order has

been changed so that the traditional scales (3 validity and 10 clinical) can be derived from the first 370 items. The remaining 197 items (371 to 567) provide different supplementary, content, and research measures. A number of new scales were included along with new, subtle, adjunctive measures of test validity, separate measures of masculinity and femininity, and 15 additional content scales measuring specific personality factors (e.g., Anxiety, Health concerns, Cynicism). An extensive research base has accumulated related to areas such as the validity of MMPI/MMPI-2 code types, use with special populations, the ability to distinguish over- or underreporting of symptoms, and comparability among the original MMPI, MMPI-2, and MMPI-A.

In 2008, Ben-Porath and Tellegen published the MMPI-2-Restructured Form (MMPI-2-RF, 2008/2011) in response to two major ongoing criticisms of the MMPI-2. (There are other criticisms, but these two are major ones the MMPI-2-RF addresses.) These criticisms were that the scales are too heterogeneous and overlapping and that the test is simply too long. Work progressed to refine the basic clinical scales by extracting the common or shared variable of demoralization. The result is the shorter and more homogeneous Restructured Clinical (RC) scales (Tellegen et al., 2003), used principally on the MMPI-2-RF. Further, five core personality scales related to psychopathology were developed (Personality Psychopathology Five [PSY-5]; Harkness, McNulty, Ben-Porath, & Graham, 2002). These and other developments have, to a certain extent, provided alternatives to the heterogeneous and more-difficult-to-interpret original clinical scales. By combining these scale groupings with the validity and additional scales, the MMPI-2-RF is a 388-item measure with some improved psychometric and interpretive qualities (Ben-Porath & Tellegen, 2008/2011). Additionally, a short form for the MMPI-2 is available through computer-adapted assessment (MMPI-2-CA; Forbey & Ben-Porath, 2007), in which future items are selected based on responses to past items. Currently, the clinical and research community is somewhat split between those who adhere to the MMPI-2 and those who have opted for the MMPI-2-RF (few actively use the MMPI-2-CA).

Early on, it was noticed that the original MMPI produced different scale elevations for adolescents than for adults. This recognition resulted in the development of different sets of recommended norms for use with adolescent populations (Archer, 1987; Colligan & Offord, 1989; Klinefelter, Pancoast, Archer, & Pruitt, 1990; Marks, Seeman, & Haller, 1974). However, many practitioners and researchers felt that, even with the use of adolescent norms, there were still considerable difficulties. Specifically, the test was too long, the reading level was too high, there was a need for contemporary norms, more of the content needed to assess problems specifically related to adolescents, and some of the language was outmoded and/or inappropriate (Archer, Maruish, Imhof, & Piotrowski, 1991). In response to these issues, the restandardization committee for the MMPI-2 decided in 1989 to develop the MMPI-Adolescent (MMPI-A), which was first made available in 1992 (Butcher et al., 1992). It was normed against a generally representative group of 805 males and 815 females between the ages of 14 and 18. The main discrepancy between the normative group and comparison with U.S. census data was that the parents of the normative group were better educated. Despite the similarity with the MMPI and MMPI-2, there are several important differences. Fifty-eight items were deleted from the original standard scales, some of the

wording of items was changed, and new items relevant to adolescent concerns were included. The result is the inclusion of four new validity scales (VRIN, TRIN, *F1*, *F2*), in addition to the earlier validity scales (*L*, *F*, *K*). There are also six supplementary scales (e.g., Immaturity Scale, Anxiety, Repression) and additional newly developed content scales (e.g., A-dep/Adolescent Depression). To counter claims that the MMPI is too long, especially for adolescents, the MMPI-A contains 478 items, thereby shortening the administration time. This can be shortened even further by administering only the first 350 items, still sufficient to obtain the validity and standard clinical scales. Thus, the MMPI-A is strongly related to the MMPI and MMPI-2 (and their respective databases) but also has a number of important distinctive features of its own.

RELIABILITY AND VALIDITY

Reliability studies on the original MMPI indicate that it had moderate levels of temporal stability and internal consistency. For example, Hunsley, Hanson, and Parker (1988) completed a meta-analysis of studies performed on the MMPI between 1970 and 1981 and concluded, "All MMPI scales are quite reliable, with values that range from a low of .71 (Scale *Ma*) to a high of .84 (Scale *Pt*)" (p. 45). Their analysis was derived from studies that included a wide range of populations, intervals that ranged from 1 day to 2 years, and a combined sample size exceeding 5,000. In contrast to Hunsley et al., some authors have reported that the fluctuations in some of the scales are sufficiently wide to question their reliabilities (Hathaway & Monachesi, 1963; Mauger, 1972). Proponents of the MMPI counter that some fluctuation in test scores is to be expected. This is especially true for psychiatric populations because the effects of treatment or stabilization in a temporary crisis are likely to be reflected in a patient's test performance (J. Graham, Smith, & Schwartz, 1986). Bergin (1971) demonstrated that Scale 2 (Depression) is particularly likely to be lowered after successful treatment. Similarly, Scale 7 (Psychasthenia) would be likely to alter according to a person's external situation and current psychological well-being. Thus, test-retest reliability may actually be a less appropriate method of evaluating these scales for certain populations. This defense of the test's reliability is somewhat undermined by the observation that test-retest reliability is actually slightly more stable for psychiatric populations than for normals. Whereas the median range for psychiatric patients is about .80, median reliabilities for normals are about .70. Split-half reliabilities are likewise moderate, having an extremely wide range from .05 to .96, with median correlations in the .70s (Hunsley et al., 1988).

Reliability reported in the MMPI-2 manual (Butcher et al., 1989) indicates moderate test-retest reliabilities. However, test-retest reliabilities were calculated for a narrow population over short-term retesting intervals. Reliabilities for normal males over an average interval of 8.58 days (*Mdn* = 57 days) ranged from a low of .67 for Scale 6 to a high of .92 for Scale 0 (Butcher et al., 1989). A parallel sample of females over the same retesting interval produced similar reliabilities ranging from .58 (Scale 6) to .91 (Scale 0). Standard error of measurements for the different scales ranged from 2 to 3 raw score points (Butcher et al., 1989, 2001; Munley, 1991).

Reliability reported in the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008/2011) indicates adequate internal consistency and moderate test-retest reliabilities. Outside of the validity scales, for the normative sample, internal consistency across all scales averaged .66, ranging in the normative sample from .38 for the SUI Scale to .88 for the RCd Scale and ranging in an outpatient sample from .56 for the BRF Scale to .94 for the EID Scale. Like the MMPI-2, the test-retest interval was very short term (1 week). Reliabilities ranged from a low of .54 for the NUC Scale to a high of .93 for the DISC-r Scale. The average test-retest reliability was .79. Standard error of measurements for the different scales ranged from 3 to 7 raw score points (Tellegen & Ben-Porath, 2008/2011).

One difficulty with the MMPI-2 lies in the construction of the scales themselves. The intercorrelations between many of the scales are quite high, which results primarily from the extensive degree of item overlap. Sometimes the same item is used simultaneously for the scoring of several different scales, and most of the scales have a relatively high proportion of items common to other scales. For example, Scales 7 (Psychasthenia) and 8 (Schizophrenia) have high overlap, which is reflected in correlations ranging from .64 to .87, depending on the population sampled (Butcher et al., 1989; Dahlstrom & Welsh, 1960). Scale 8, which has the highest number of items (78), has only 16 items that are unique to it (Dahlstrom, Welsh, & Dahlstrom, 1972). Similarly, Scale *F* (Infrequency) is highly correlated with Scales 7 (*Pt*), 8 (*Sc*), and the Bizarre Mentation content scale. The practical implication is that interpreters need to be quite cautious about inferring a “fake bad” profile if profile *F* is elevated along with 7 (*Pt*), 8 (*Sc*), and Bizarre Mentation. Several factor-analytic studies have been conducted that were motivated in part by a need to further understand the high intercorrelations among scales. These studies have not found any consistent numbers and types of factors. The numbers of factors range between 2 (Dahlstrom, Welsh, & Dahlstrom, 1975; Dahlstrom et al., 1972; D. Jackson, Fraboni, & Helms, 1997) and 9 (Archer & Krishnamurthy, 1997a; Costa, Zonderman, Williams, & McCrae, 1985) and even up to 21 (J. H. Johnson, Null, Butcher, & Johnson, 1984). This finding suggests that these factors are not highly differentiated.

The different MMPI-2 scales correlate so highly in part because the original selection of the items for inclusion in each scale was based on a comparison of normals with different clinical groups, separately for each scale. The items, then, were selected based on their differentiation of normals from various psychiatric populations, rather than on their differentiation of one psychiatric population from another. Although the psychiatric groups varied from the normals on several traits, this manner of scale construction did not develop accurate measurements of these different traits. Rather, the scales are filled with many heterogeneous items and measure multidimensional, often poorly defined attributes. This approach has also led to many items being shared with other scales. In contrast, an approach in which specific psychiatric groups had been compared with one another would have been more likely to have resulted in scales with less item overlap and with the ability to measure more unidimensional traits.

A partial defense of item overlap is that for complex, multidimensional variables such as pathological syndromes, important relationships would be expected with other similar constructs. If these other constructs were being measured on the same test, it would further be expected that there would be scale overlap on these theoretically

and clinically related syndromes (Dahlstrom et al., 1972). For example, depression is a common feature among several categories of psychopathology. Thus, it would be theoretically related to conditions such as hypochondriasis, schizophrenia, and anxiety. This means that it would be expected that the common occurrence of depression would result in intercorrelations between scales and would produce scales that, while intercorrelated, would still have subtle and clinically different meanings (Broughton, 1984). Thus, the multidimensionality of the scales, combined with their item overlap, would be not so much a weakness of the MMPI-2/MMPI-A but would be expected, given the nature of the constructs. Accurate interpretation, however, would need to include an awareness of the subtle differences and similarities between scales.

Despite this defense, differentiation among different discrete psychopathological presentations is important, and the MMPI-2-RF altered the structure of its scales to account for this. With small exceptions of a few validity scales, items on individual scales of the MMPI-2-RF contribute only to the calculation of those individual scales, with no item loading onto multiple scales. This allows for “purer” interpretation of each of the scales, as elevation in one scale cannot be the result of elevation in another. Although this may not mirror the reality of how psychopathology functions—that is, rarely do disorders function completely independently of one another—it does allow for more confidence in differential diagnosis.

An issue related to MMPI-2/MMPI-A scale multidimensionality is that elevations can often occur for a variety of reasons. For example, an elevation on 4 (Psychopathic Deviate) might result from family discord, poor peer relations, alienation from self and society, and/or acting out associated with legal difficulties. A person interpreting an elevated Scale 4 (Psychopathic Deviate) might potentially infer antisocial acting out when family discord is the major reason for the scale elevation. To enhance the likelihood of accurate interpretations, practitioners need to carefully evaluate the meanings of scale elevations. This might include looking at the content of selected items (critical items), scoring the Harris-Lingoes subscales, considering the meanings of content or supplementary scales, reviewing scores on the restructured clinical scales, referring to published MMPI research, and integrating the results from the client’s history and relevant behavioral observations. Differentiating which of these scale dimensions is most relevant can be quite challenging for the practitioner. This is not as significant a problem on the MMPI-2-RF, which has scales that are much more streamlined (many fewer items) and as a result is significantly more homogeneous in content. Elevation on RC4 (Antisocial Behavior) is much more directly representative of, and thus interpretable as, antisocial behavior, as fewer items load onto the scale, and the items that do contribute to the scale are much more similar and related in content with each other.

The difficulties associated with reliability and scale construction have led to challenges to the MMPI’s validity. Rodgers (1972) even referred to the MMPI as a “psychometric nightmare.” However, although the psychometric difficulties have presented problems, these problems have been somewhat compensated for by extensive validity studies. More specifically, the meanings of 2- and 3-point MMPI/MMPI-2 profile code types have been researched extensively, as have the contributions that the MMPI can make toward assessing and predicting specific problem areas. There are at least 8,000 studies investigating profile patterns, and this number is continually increasing (e.g., Butcher, 2011; DuAlba & Scott, 1993; Gallucci, 1994; J. Graham, Ben-Porath,

& McNulty, 1999; McNulty, Ben-Porath, & Graham, 1998). These studies provide extensive evidence of the MMPI's construct validity. For example, elevations on Scales 4 (*Pd*) and 9 (*Ma*) have been associated with measures of impulsivity, aggression, substance abuse, and sensation seeking among adolescent inpatients (Gallucci, 1994). In addition, the degree to which individuals improve from psychotherapy was predicted based on elevations on the content scales of Anxiety (*ANX*) and Depression (*DEP*; Chisholm, Crowther, & Ben-Porath, 1997). Finally, high scores on Scale 0 (*Si*) have been associated with persons who have low self-esteem, social anxiety, and low sociability (Sieber & Meyers, 1992). Individual clinicians can consult research on code types to obtain specific personality descriptions and learn of potential problems to which a client may be susceptible. The extensiveness and strength of these validity studies have usually been regarded as major assets of the MMPI and are important reasons for its continued popularity.

Because it is newer, the MMPI-2-RF has less research supporting its validity, though the research to date is significant and convincing. The substantive scales have been found to be strongly associated with appropriate ratings of mental health professionals after an intake interview or a few sessions for an outpatient population (J. R. Graham, Ben-Porath, & McNulty, 1999), external ratings for a psychiatric inpatient sample (Arbisi, Ben-Porath, & McNulty, 2003), self-report ratings for a veteran sample, forensic samples, and a college sample (Tellegen & Ben-Porath, 2008/2011). The MMPI-2-RF scales are also highly correlated with the expected MMPI-2 scales. Intercorrelations among the substantive scales on the MMPI-2-RF relate clearly and directly to the interpretive strategies recommended, with the highest correlations found among scales with similarly themed content (e.g., thought scales; behavior scales; Tellegen & Ben-Porath, 2008/2011).

In addition to studying the correlates of code type, another approach to establishing validity is to assess the accuracy of inferences based on the MMPI. Early studies by Kostlan (1954) and Little and Shneidman (1959) indicated that the MMPI was relatively more accurate than other standard assessment instruments, especially when the MMPI was combined with social case history data. This incremental validity of the MMPI has been supported in later reviews by Garb (1998) and J. R. Graham and Lilly (1984). For example, the accuracy of neurologists' diagnoses was found to increase when they added an MMPI to their patient data (S. Schwartz & Wiedel, 1981). Garb (1998b) concluded that the MMPI was more accurate than social history alone and was superior to projectives, and that the highest incremental validity was obtained when the MMPI was combined with social history. In addition, incremental validity of the new MMPI-2 content scales has been found in that they both expanded on and increased the validity of the standard clinical scales (Barthlow, Graham, Ben-Porath, & McNulty, 1999; Ben-Porath, McCully, & Almagor, 1993).

The MMPI-2-RF restructured clinical scales have been found to have better discriminant validity than the basic clinical scales (Sellbom, Ben-Porath, & Graham, 2006). Many of the scales have been found to be superior to associated MMPI-2 scales at predicting different outcomes, and many have been found to have incremental validity, making significantly better predictions than without using the MMPI-2-RF. For example, the Response Bias Scale (RBS) better predicted overreporting of memory complaints than the MMPI-2 F, Fb, Fp, and FBS scales (Gervais, Ben-Porath, Wygant,

& Sellbom, 2010). Additionally, the RCd Scale and other scales related to somatoform and interpersonal problems added predictive power to presurgical screening of spine surgery patients, focused on likelihood of continued pain and dysfunction after physical recovery (Marek, Block, & Ben-Porath, 2015). The measure overall has been found useful in bariatric surgery evaluations (Marek et al., 2013) and predicting premature termination (Anestis, Finn, Gottfried, Arbisi, & Joiner, 2015), negative treatment outcomes (Anestis, Gottfried, & Joiner, 2015), and malingering (Goodwin, Sellbom, & Arbisi, 2013; Sellbom, Toomey, Wygant, Kucharski, & Duncan, 2010; Sellbom, Wygant, & Bagby, 2012; Whitney, Davis, Shepard, & Herman, 2008). In general, the Restructured Clinical scales have also shown incremental validity in predicting clinical symptoms (Sellbom, Graham, & Schenk, 2006).

ASSETS AND LIMITATIONS

The previous discussion on reliability and validity highlights several issues associated with the MMPI-2. These include moderate levels of reliability, extensive length, and problems related to the construction of the scales, such as item overlap, high intercorrelations among scales, and multidimensional and somewhat poorly defined variables. Some of the older criticisms of the original MMPI relating to obsolete norms, offensive items, and poorly worded items have been largely corrected with the publication of the MMPI-2 and MMPI-A. The MMPI-2 also has a number of strengths along with other weaknesses. The overwhelming majority of these criticisms were also addressed in the development of the MMPI-2-RF, though a major limitation of the MMPI-2-RF is that the extremely extensive history of research on the MMPI and MMPI-2 are not applicable to the new test. As such, a rich history of compelling empirical work, including much about the clinical utility of the MMPI and MMPI-2, is not applicable to the new measure.

One caution stemming from the construction of the original MMPI is that it generally does not provide much information related to normal populations. The items were selected on the basis of their ability to differentiate a bimodal population of normals from psychiatric patients. Thus, extreme scores can be interpreted with a high degree of confidence, but moderate elevations must be interpreted with appropriate caution. An elevation in the range of 1 standard deviation above the mean is more likely to represent an insignificant fluctuation of a normal population than would be the case if a normally distributed group had been used for the scale construction. This is in contrast to a test such as the NEO Personality Inventory, which used a more evenly distributed sample (as opposed to a bimodal one) and, as a result, can make meaningful interpretations based on moderate elevations. The MMPI-2 and MMPI-2-RF partly address this difficulty as they have used broad contemporary norms for their comparisons, combined with uniform *T* scores (Butcher et al., 2001; Tellegen & Ben-Porath, 2008/2011). However, evaluation of normals can be complicated by the observation that normal persons sometimes achieve high scores. Despite these difficulties, the use and understanding of nonclinical populations have been increasing (Keiller & Graham, 1993). In particular, uses have included screening personnel for sensitive jobs, such as air traffic controllers, police officers, and nuclear plant operators.

Although there are great similarities between the MMPI and MMPI-2, issues have been raised regarding comparability between the two versions, raising questions about just how applicable MMPI research is to the MMPI-2. In defense of their comparability are the many similarities in format, scale descriptions, and items. In particular, Ben-Porath and Butcher (1989) found that the effects of rewriting 82 of the MMPI items for inclusion in the MMPI-2 were minimal. The rewritten items had no effect on any of the validity, clinical, or special scales when comparisons were made between administrations of the original and restandardized versions using college students. This finding provided some support for Butcher and Pope's (1989) contention that the MMPI-2 validity and clinical scales measure "exactly what they have always measured" (p. 11). Further studies have generally found that there are few differences based on individual scale comparisons (Ben-Porath & Butcher, 1989; Chojnacki & Walsh, 1992; Harrell, Honaker, & Parnell, 1992; L. Ward, 1991). Similarly, number of elevated scales between the two forms does not seem to be significantly different, and there has been 75% agreement regarding whether a subject's profile was considered to be within normal limits (Ben-Porath & Butcher, 1989).

Despite these similarities, the use of the restandardization norms and the use of uniform *T* scores have created differences in 2-point codes among different population samples, including differences among 31% of the code types derived from general psychiatric patients (Butcher et al., 1989), 22% of peace officers (Hargrave, Hiatt, Ogard, & Karr, 1994), 39% to 42% of psychiatric inpatients (D. Edwards, Morrison, & Weissman, 1993), and a full 50% of both university students and forensic populations (Humphrey & Dahlstrom, 1995). The greatest level of disagreements was for poorly defined code types (mild to moderate elevations combined with more than two "competing" scales). In contrast, well-defined code types (highly elevated and without "competing" third- or fourth-most elevated scales) had considerably higher concordance (Tellegen & Ben-Porath, 1993). This fact suggests that special care should be taken regarding poorly defined code types, and, if more than two scales are elevated, the meanings of the relatively high scales not included in the code should be given particular interpretive attention.

These discrepancies in code types seem to question the exact transferability of past code type research on the MMPI onto the more recent MMPI-2 (and MMPI-A). However, the most important question is the extent to which the MMPI-2 accurately describes an individual's relevant behaviors. The research that has been done on the MMPI-2 does support the conclusion that scores on the MMPI-2 predict the same sorts of behaviors that were found with the earlier MMPI (e.g., Archer, Griffin, & Aiduk, 1995; Butcher et al., 2001; J. R. Graham et al., 1999; Timbrook & Graham, 1994). This is also true on the stand-alone research on the MMPI-2-RF (e.g., Haber & Baum, 2014; Rock, Sellbom, Ben-Porath, & Salekin, 2013; Sellbom, Ben-Porath, Baum, Erez, & Gregory, 2008; Tarescavage, Luna-Jones, & Ben-Porath, 2014).

As highlighted in the previous section, a traditional asset of the MMPI/MMPI-2/MMPI-A has been extensive and ongoing code type studies. However, difficulties with these studies have been noted. First, some studies have tried to be extremely inclusive in deciding which codes to evaluate. In contrast, others have been quite restrictive (i.e., including only clearly defined code types). Inclusion/exclusion among the different studies has ranged from 24% to 99% (McGrath & Ingersoll, 1999a). The practical

implication for clinicians is considering the degree to which their code type classifications parallel those of research. If specific clinicians are highly inclusive about what they consider to be interpretable code types, they may place unwarranted faith in their interpretations if the body of research they are drawing from has used quite restrictive criteria (e.g., J. R. Graham et al., 1999, used only well-defined code types). A further concern is that the mean effect size across studies was quite variable, with a high of .74 and low of .02 (McGrath & Ingersoll, 1999b; Meyer & Archer, 2001). In addition, effect sizes were found to vary among different scales and code types. Therefore, not only may practitioners be placing unwarranted faith in some of their interpretations, but the validity of the interpretations they do make is likely to vary according to which scale/code type they are interpreting.

The MMPI-2 scale labels can be misleading because they use traditional diagnostic categories, while the MMPI-2-RF worked hard to make the scale labels more directly interpretable. A person might read a scale such as schizophrenia and infer that a person with a peak on that scale fits the diagnosis of schizophrenia. Although it was originally hoped that the MMPI could be used to make differential psychiatric diagnoses, it was soon found that the MMPI could not perform this function adequately. Thus, even though individuals with schizophrenia score high on Scale 8, so do other psychotic and nonpsychotic groups. Also, moderate elevations can occur for some normal persons. With each progressive edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM; American Psychiatric Association [APA] 1968, 1980, 1987, 1994, 2000, 2013), the labels given to the scale names have become progressively more outdated. This potentially causes confusion related to diagnosis because the scales reflect these outdated categories. For example, Scales 1, 2, and 3 are called the *neurotic triad*, and Scale 7 is labeled *Psychasthenia*; yet clinicians are often faced with the need to translate these outdated designations into DSM-5 (APA, 2013) terminology. This difficulty has been somewhat alleviated through research detailing the frequencies which various diagnoses derived from recent editions of the DSM occur on the different code types (Bagby et al., 2005; Morey, Blashfield, Webb, & Jewell, 1988; Vincent et al., 1983). DSM translations have been further aided through the use of different content, supplementary, and restructured scales that allow for broader descriptions of symptom patterns (Barthlow et al., 1999; Butcher, 2006, 2011; Butcher et al., 1990; Graham, 2011; C. L. Williams et al., 1992). Again, the MMPI-2-RF has addressed this issue by naming the scales more closely to their actual content and interpretive directions.

To compensate for the difficulties related to MMPI-2 scale labels, clinicians should become aware of the current meanings of the scales based on research rather than the meanings implied by the often-misleading scale titles. This approach can be aided in part by using scale numbers rather than titles. For example, Scale 8 suggests attributes such as apathy, feelings of alienation, philosophical interests, poor family relations, and unusual thought processes rather than schizophrenia. It is the clinician's responsibility to determine which of these attributes are most characteristic of the person being evaluated. Clinicians should also be aware of the relationships among scales as represented by the extensive research performed on 2- and 3-point code types. Usually the patterns or profiles of the scales on the MMPI-2 are far more useful and valid than merely considering individual scale elevations. The extensive research in this area represents what is probably the strongest asset of the MMPI-2. This volume of work has

prevented the MMPI from becoming obsolete, despite outdated scale names, and has been instrumental in transforming it from a test of psychiatric classification into a far more wide-band personality inventory.

A further significant asset is the MMPI-2's immense popularity and familiarity within the field. Extensive research has been performed in a variety of areas, and new developments have included abbreviated forms, new scales, the use of critical items, an adolescent version, and computerized interpretation systems. The MMPI has been translated into more than 50 languages and is available in numerous countries. Normative and validity studies have been conducted on several different cultural groups (see Butcher, 1996, 2004; Handel & Ben-Porath, 2000), which makes possible the comparison of data collected from varying cultures. In contexts where no norms have been developed, at least the test format lends itself to the development of more appropriate norms that can then be used in these contexts. These strengths contrast with the newer MMPI-2-RF, which, although it already has a strong research base, has not yet had the opportunity to become as widespread as the MMPI-2.

A complicating aspect of the MMPI-2 and MMPI-2-RF is that interpretations often need to take into account many demographic variables (e.g., Schinka, LaLone, & Greene, 1998). It has been demonstrated that age, sex, race, place of residence, intelligence, education, and socioeconomic status are all related to the MMPI scales. Often the same relative elevation of profiles can have quite different meanings when corrections are made for demographic variables, especially on the MMPI-2. Some of the more important and well researched of these are discussed later in the chapter (see the section titled "Use with Diverse Groups").

The advantages and cautions for using the MMPI-2 and MMPI-A indicate that a considerable degree of psychological sophistication by clinicians is necessary, which is also true but less important for the MMPI-2-RF. Both their assets and limitations need to be understood and taken into account. The limitations for the original MMPI are numerous and include moderately adequate reliability, heterogeneity of the clinical scales, offensive items, limited usefulness for normal populations, misleading labels for the scales, and excessive length. However, the limitations of the MMPI-2 are balanced by a number of significant assets. For example, the excessive length can be countered either by administering only the first 370 items (or the first 350 items for the MMPI-A). An additional important strength is the extensive research relating to the meanings of the different scales and the relationships among scales. Extensive strategies are also in place to help refine and expand the meanings of scale elevations by using alternative scales (content, Harris-Lingoes, supplementary). Further assets are the MMPI-2's familiarity in the field, the development of subgroup norms, and extensive research in specific problem areas. The MMPI-2-RF continues to gain popularity in the field based on strong psychometric properties and clinical utility. Of central importance is that the MMPI has repeatedly been proven to have practical value for clinicians, especially because the variables that the scales attempt to measure are meaningful and even essential areas of clinical information. The over 10,000 studies on or using it, combined with its extensive clinical use, provide ample evidence of its popularity.

USE WITH DIVERSE GROUPS

Age

A significant feature of adolescent populations is a general elevation on many of the original MMPI and MMPI-2 scales. This was particularly true for scales related to level of energy (9) and rebellious acting out (4). These generally higher elevations have led to considerable controversy over whether adolescents have more actual pathology (based on external behavioral correlates) or whether they merely have higher scores without correspondingly higher pathology (Archer, 1984, 1987, 1992a; Janus, Tolbert, Calestro, & Toepfer, 1996). The controversy has encouraged efforts to more clearly understand behavioral correlates of adolescent profiles (Archer, 2005; Archer & Jacobson, 1993; Basham, 1992; Bolinsky, Trumbetta, Hanson, & Gottesman, 2010; Janus et al., 1996). The controversy has also led to the development of the MMPI-A. The general consensus seems to be that using the MMPI-A (and norms based on it) results in behavioral descriptions that are at least as accurate as descriptors based on the MMPI/MMPI-2 (Archer, 1992a, 1992b; Butcher et al., 1992; Janus et al., 1996; Weed, Butcher, & Williams, 1994). In fact, when the MMPI-A was given to a sample of 18-year-olds instead of the MMPI-2, the MMPI-A provided generally lower clinical scale values than the MMPI-2, suggesting that using the MMPI-A may better normalize the behaviors of adolescence (Shaevel & Archer, 1996).

When assessing 18-year-olds, clinicians must decide whether to use the MMPI-2 or the MMPI-A. If the adolescents are living independently and are relative mature, clinicians should consider using the MMPI-2 (J. R. Graham, 2011). In contrast, if they are still living at home and are relatively immature, the MMPI-A is recommended.

As people age, they generally have reduced energy and greater focus on health concerns. Whereas this trend is reflected in scores on the MMPI-2 of older adults, these changes tend to be fairly small (less than 5 *T* scores) and clinically nonsignificant (J. R. Graham, 2011). As a result, using separate norms for older adults has not been recommended.

Although the MMPI-2-RF does not yet have an adolescent version, there is some suggestive evidence that the constructs from the MMPI-2-RF will apply to adolescent samples (e.g., Trumbetta, Bolinsky, & Gottesman, 2013).

Ethnicity

The MMPI/MMPI-2 has been studied extensively to determine how appropriate it is to use with culturally divergent groups. This research has centered on both ethnically different (minority) groups within the United States and use in different countries. Although very little research has focused on evaluating ethnic and cultural differences on the MMPI-2-RF, the literature and conclusions discussed here can be applied tentatively to the MMPI-2-RF. There are a wide variety of possible reasons why persons from different cultural groups might score in a certain direction. Although scores may

be due to the accurate measurement of different personality traits, they may also be the result of cultural tendencies to acquiesce by giving socially desirable responses, differing beliefs about modesty, role conflicts, or varying interpretations of the meaning of items. Profiles may also reflect the results of racial discrimination in that scales associated with anger, impulsiveness, and frustration may be elevated.

MMPI/MMPI-2 research on ethnic groups within the United States has centered on differences between African Americans versus White individuals. Research on African American versus White individuals' MMPI performance has frequently indicated that African Americans are more likely to score higher on Scales *F*, 8, and 9 (Green & Kelley, 1988; Gynther & Green, 1980; C. Smith & Graham, 1981). This finding has resulted in considerable controversy over whether these differences indicate higher levels of actual pathology or merely reflect differences in perceptions and values without implying greater maladjustment. If the differences do not reflect greater actual pathology, then specialized subgroup norms are required to correct for this source of error. However, reviews of over 30 years of research have concluded that, although African versus European American differences could be found for some populations, there was no consistent pattern to these differences across all populations (Greene, 1987, 1991; G. C. N. Hall et al., 1999; Knaster & Micucci, 2013). What seemed of greater significance was the role of moderator variables, such as education, income, age, and type of pathology. When African American and White psychiatric patients were compared according to level of education and type of pathology, their MMPI/MMPI-2 performances were the same (McNulty, Graham, Ben-Porath, & Stein, 1997; Timbrook & Graham, 1994). In other words, behavioral correlates between African American and White individuals' MMPI performance have generally not found differences between the two groups. For example, ratings by clinicians (McNulty, Graham, Ben Porath, & Stein, 1997) and partners (Timbrook & Graham, 1994) were equally as accurate for both groups. In addition, the main behavioral features of 68/86 code types between African Americans and European Americans were the same (Clark & Miller, 1971). Furthermore, predictions based on African American and White juvenile delinquents' MMPI scores were equally accurate (Green & Kelley, 1988; Timbrook & Graham, 1994). A final crucial finding has been that, even when mean differences have been found, they have been of less than 5 *T* score points (G. C. N. Hall et al., 1999; Stukenberg, Brady, & Klinetob, 2000). The magnitude of this difference is not clinically meaningful. Based on these findings, it would be premature to develop and use separate norms for African Americans. However, it still is important for clinicians to continually be aware of any possible culturally relevant factors (e.g., effects of discrimination) that may cause unique elevations in an individual African American's profile.

Native Americans tend to score higher on a number of MMPI-2 scales (including *L*, *F*, *K*, 1, 4, 6, 7, 8, 9, *HEA*, *BIZ*, *CYN*, *ASP*, and *TRT*; Pace et al., 2006; Robin, Greene, Albaugh, Caldwell, & Goldman, 2003). While this may be the result of pathologizing the general worldviews of this population, as opposed to accurately assessing pathology (Hill, Pace, & Robbins, 2010), further research found that when matching for age, gender, and education, the cultural differences in scores disappeared, suggesting that adverse conditions account for the differences seen in the scores of Native Americans rather than test bias (Robin et al., 2003). As a result, low elevations should

be interpreted with caution. In contrast, once scores become higher than $T = 65$, the elevations are likely to reflect actual psychopathology.

Similar to African American versus White comparisons, no consistent patterns have been found across different populations for Latino/Hispanic and Asian Americans. Differences between Latino Americans and White counterparts have generally been found to be less than African American or European American differences (Greene, 1991). The largest difference was that male Latinos scored higher on Scale 5 than male European Americans (G. C. N. Hall et al., 1999). However, all differences were still less than 5 T score points (G. C. N. Hall et al., 1999), and other samples did not find significant group differences (Whitworth & McBlaine, 1993). Similarly, studies comparing Asian American and White samples have found little to no significant group differences (Tsushima & Tsushima, 2009). Given the reviews of ethnicity and the MMPI/MMPI-2 (Butcher, 2004; J. R. Graham, 2011; Greene, 1987, 1991; G. C. N. Hall et al., 1999; Schinka et al., 1998), these conclusions seem warranted (again, although based on the MMPI/MMPI-2, high correlations with the MMPI-2-RF suggest that these suggestions should be tentatively followed for the MMPI-2-RF as well):

- Even when ethnic differences have been found between various groups, overall these differences are less than 5 T score points (less than 10% of the variance), and they are therefore not clinically meaningful.
- It would be premature to develop new norms for ethnic groups, particularly since moderator variables (socioeconomic status, education, age) seem to explain most of the variance in performance.
- It may at times be useful to consider the meanings of ethnic score differences for specific ethnic subgroups. For example, Latino workers' compensation cases may be more likely to somatize psychological distress as reflected by greater elevations on 1 (H_s), 2 (D), and 3 (H_y) than White Americans (DuAlba & Scott, 1993). In addition, higher CYN (Cynicism) and ASP (Antisocial Practices) found among African American as opposed to White forensic populations are likely to represent clinically meaningful differences (Ben-Porath, Shondrick, & Stafford, 1995).
- Low scores for Native Americans are likely to reflect cultural factors rather than psychopathology, but higher elevations (above $T = 65$) are likely to reflect actual psychopathology.
- Future research should consider within-group ethnic differences, including degree of identification with his or her ethnic group, acculturation (e.g., Lessenger, 1997; Tsai & Pike, 2000), language fluency, perceived minority status, and degree to which the person feels discriminated against.
- More research needs to investigate the relationship between ethnicity and the many supplementary and content scales.

The MMPI/MMPI-2 has been used not only with multiple ethnic groups within the United States; it has also been used in a wide variety of different countries. An important rationale for this use is that it is more efficient to adapt and validate the MMPI/MMPI-2 for a different country than go to the far more extensive effort of developing a whole new test for the culture. Examples of countries where adaptations

have occurred include such diverse areas as China, Israel, Pakistan, South Africa, Chile, Mexico, and Japan (see Butcher, 1996). Whenever clinicians work with different cross-national groups, they should consult the specific norms that have been developed for use with these groups as well as become familiar with any research that may have been carried out with the MMPI on these groups. Useful sources are Butcher's (1996) *International Adaptations of the MMPI-2* and reviews of cross-cultural research by Greene (1987, 1991) and G. C. N. Hall et al. (1999). In general, either no or small, non-clinically meaningful differences have been found for samples from different countries, including Korea (Han et al., 2013; S. Kim, Goodman, Toruno, Sherry, & Kim, 2015), China (Kwan, 1999), Cuba (Quevedo & Butcher, 2005), Vietnam (Dong & Church, 2003), and Israel (Shkalim, 2015). Small differences were found in samples in some countries, such as Israel (Almagor & Koren, 2001) and Brunei (Mundia, 2011), but again these differences were not significantly meaningful clinically. Early research has revealed strong psychometric properties for the MMPI-2-RF in Israel (Shkalim, 2015), but very little other cross-cultural research has been done with the measure.

ADMINISTRATION

The MMPI-2 and MMPI-2-RF can be administered to persons who are 16 years of age or older with an eighth-grade reading level. As noted, it is possible to administer the MMPI-2 to persons between the ages of 16 and 18, but adolescent norms need to be used. However, the preferred option for individuals between ages 14 and 18 is to have them take the MMPI-A. It is often helpful to augment the standard instructions in the MMPI-2, MMPI-2-RF, and MMPI-A booklets. In particular, examiners should explain to clients the reason for testing and how the results will be used. It might also be pointed out that the test was designed to determine whether someone has presented him- or herself in an either unrealistically positive or exaggeratedly disturbed manner. Thus, the best strategy is to request that examinees be as honest and as clear as possible. Finally, it might be clarified that some, or even many, of the questions might seem a bit unusual. They have been developed to assess individuals with a wide range of personality styles and problems. If they do not apply to the person taking the test, this should be indicated with either a true or false response. Including this additional information is likely to result in less anxiety, more accurate responses, and greater rapport. Completion times for all persons taking the test should be noted.

Completion of the first 370 items on the MMPI-2 and first 350 items on the MMPI-A allows for the scoring of the basic validity and standard clinical scales, while the MMPI-2-RF must be taken in its entirety. The final 197 MMPI-2 and 128 MMPI-A items are used for scoring different supplementary and content scales. An online computer administration is available through Pearson. For persons who have special difficulties, the MMPI-2 has available an individual (Box) form and a tape-recorded form. The Box form is most appropriate for persons who have difficulties concentrating and/or reading. Each item is presented on a card, which the person is requested to place into one of three different sections to indicate a "true," "false," or "cannot say" response. The tape-recorded form is used for persons who have reading difficulties because of factors such as illiteracy, blindness, or neurological reasons.

Some clinicians allow the client to take the MMPI under unsupervised conditions (such as at home). Butcher and Pope (1989) stressed that this is not recommended, for the following reasons:

- The conditions are too dissimilar from those used for the normative samples and any significant change in proceedings might alter the results.
- Clients might consult others to determine which answers to make.
- The clinician cannot be aware of possible conditions that might compromise reliability and validity.
- There is no assurance that the client will actually complete the protocol him- or herself.

Thus, any administration should closely follow the administration procedures used for the normative samples. This means providing clear, consistent instructions; ensuring that the directions are understood; providing adequate supervision; and making sure the setting will enhance concentration by limiting noise and potential interruptions.

MMPI-2 INTERPRETATION PROCEDURE

The next eight steps are recommended for interpreting MMPI-2/MMPI-A profiles. These steps should be followed with a knowledge and awareness of the implications of age, culture, intellectual level, education, and level of functioning as well as the reason, motivation, and context of assessment. While looking at the overall configuration of the test (Steps 4, 5, and 6), clinicians can elaborate on the meanings of the different scales and the relationships among scales by consulting the interpretive hypotheses associated with them. These can be found in later sections of this chapter on validity scales, clinical scales, and 2-point codes, as well as in sections on the content, supplementary, and other scales. The discussion of the various scales and codes represents an integration and summary of both primary sources and these MMPI-2/MMPI-A resources: Archer (2005); Butcher (2006, 2011); Butcher et al. (2001); Friedman, Lewak, Nichols, and Webb (2000); J. R. Graham (2011); Greene (2000); Greene and Clopton (1994), and Levak, Siegel, and Nichols (2011). In particular, the subsections on treatment implications have drawn on the work of Butcher (1990), Friedman et al. (2000), and Greene and Clopton (1994). Occasionally, additional quite recent material and relevant reviews/meta-analyses have been cited either to update material related to scale descriptions or to highlight important areas of research.

Step 1. Completion Time

The examiner should note the length of time required to complete the test. For a person with mild disturbance who is 16 years or older with an average IQ and eighth-grade education, the total completion time for the MMPI-2 should be approximately 90 minutes. Computer administrations are usually 15 to 30 minutes shorter (60–75 minutes in total). The MMPI-A usually takes 60 minutes to complete, with

computer administrations taking 15 minutes less time (45 minutes in total). If 2 or more hours are required for the MMPI-2 or 1.5 or more for the MMPI-A, these interpretive possibilities must be considered:

- Major psychological disturbance, particularly a severe depression or functional psychosis
- Obsessive indecision
- Below-average IQ or poor reading ability resulting from an inadequate educational background
- Cerebral impairment

If, however, an examinee finishes in less than 60 minutes, the examiner should suspect the possibility of an invalid profile, an impulsive personality, or both.

Note any erasures or pencil points on the answer sheet. The presence of a few of these signs may indicate that the person took the test seriously and reduces the likelihood of random marking; a great number of erasures may reflect obsessive-compulsive tendencies.

Step 2. Score and Plot the Profile

Complete the scoring and plot the profile. If examiners would like to score and profile the content scales, Harris-Lingoes and *Si* content subscales, the most frequently used supplementary scales, restructured clinical scales, or the personality psychopathology five scales, additional keys and profile forms may be obtained through Pearson. In addition to the possibility of scoring alternative scales, clinicians should compile further information, including IQ scores, relevant history, demographic variables, and observations derived from Step 1.

Score the critical items and note which ones indicate important trends. It is often helpful at some point to review these items with the client and obtain elaborations. In particular, it is essential to determine whether the person understood what the item was asking. Similarly, it can sometimes be helpful to examine the answer sheet and note which, if any, questions were omitted. A discussion with the client about why he or she chose not to respond might shed additional light on how he or she is functioning psychologically and what areas are creating conflict for him or her.

Step 3. Organize the Scales and Identify the Code Type

The scores can be summarized by simply listing the scores according to the order in which they appear on the profile sheet (VRIN, TRIN, *L*, *Fb*, *Fp*, *L*, *K*, *S*, 1, 2, 3, etc.) with their *T* scores to the right of these scales. For the purposes of communicating scale scores, *T* scores rather than raw scores should be used.

Developing summary codes (“code types”) provides a shorthand method for recording and communicating MMPI-2/MMPI-A results. Code types can be determined simply by looking at the two highest scale elevations. For example, the two highest scores in a profile might be 8 and 7 resulting in an 87/78 code type. The 87/78 code type can then be looked up in the “MMPI-2 Two-Point Codes” section later in this

chapter to obtain various descriptions relating to that code type. Note that Scales 5 (Masculinity-Femininity) and 0 (Social Introversion) are not strictly clinical scales, so they are not used in determining code type. Examiners should keep in mind that only well-defined code types can be interpreted safely (Butcher, 2006; D. Edwards et al., 1993; Greene, 2000; McNulty et al., 1998; Tellegen & Ben-Porath, 1993). A well-defined code type is one in which the elevated scales are above 65 and the scales used to determine the code type are 5 or more *T* score points above the next highest scales. Less well-defined profiles should be interpreted by noting each scale that is elevated and then integrating the meanings derived from the different descriptors.

Step 4. Determine Profile Validity

Assess the validity of the profile by noting the pattern of the validity scales. There are a number of indicators suggesting invalid profiles, which are described in the “MMPI-2 Validity Scales” section later in this chapter. However, the basic patterns include a defensive style in which pathology is minimized (elevated *L*, *K*, and *S* on the MMPI-2 and *L* and *K* on the MMPI-A), an exaggeration of pathology (elevated *F*, *Fb*, *Fp*, *FBS* on the MMPI-2 or *F*, *F1*, or *F2* on the MMPI-A), or an inconsistent response pattern (elevated *VRIN* or *TRIN*). In addition, clinicians should consider the context of the assessment to determine whether a defensive, fake bad, or inconsistent response style supports what is known about the client and his or her situation. In particular, the examiner should determine the likelihood that the examinee would potentially gain from over- or underreporting psychopathology.

Step 5. Determine Overall Level of Adjustment

Note the number of scales over 65 and the relative elevation of these scales. The degree to which *F* is elevated can also be an excellent indicator of the extent of pathology (assuming that it is not so high as to indicate an invalid profile). The greater the number and relative elevation of these scales, the more likely the individual is to have difficulties carrying out basic responsibilities and to experience social and personal discomfort.

Step 6. Describe Symptoms, Behaviors, and Personality Characteristics

This step represents the core process in interpretation. Mild elevations on individual scales ($T = 60-65$) represent tendencies or trends in the individual's personality. Interpretations should be treated cautiously with the more extreme descriptors being deleted or rephrased to represent milder characteristics. Scores in this range on the MMPI-A are highlighted by shading, thereby designating a marginal or transitional zone between normality and pathology. Elevations above 65 on the MMPI-2 and MMPI-A are more strongly characteristic of the individual and, with progressively greater increases, are more likely to represent core features of personality functioning. However, basing interpretations solely on specific *T* score elevations may be misleading because a client's demographic characteristics or level of functioning might alter the interpretations. For example, a high-functioning professional with a mild to moderate score on 2 (depression) probably indicates a level of introspection and

mild dysphoria that he or she is able to control effectively. In contrast, a similar elevation on a low-functioning psychiatric patient is more likely to reflect aspects of psychopathology. Furthermore, different authors use different criteria for determining high and low scores. Some authors have used *T* score ranges (e.g., $T = 70-80$); others have defined elevated scores as the upper quartile; and still others have defined a high score as the highest in a profile regardless of other *T* score elevations. As a result, the descriptors in the following sections of this chapter on interpretation do not designate specific *T* score elevations. Instead, more general descriptions associated with high and low scores have been provided. Clinicians will need to interpret the accuracy of these potential meanings by taking into consideration not merely the elevations but other relevant variables as well. In addition, each of the descriptions is modal. The descriptions should be considered as possible interpretations that will not necessarily apply to all persons having a particular score. They are merely hypotheses in need of further verification. This point is highlighted by the finding that somewhere in the range of 40% of computer-generated descriptors do not apply to the person being assessed (Butcher & Williams 2000). In addition to the Basic/Clinical Scales presented in Table 7.1, Table 7.2 presents the content scales used in the interpretation of the MMPI-2 and MMPI-A. Supplementary scales are presented later in the chapter.

Whereas *T* scores are not provided for most scale interpretations, they have been included in the subsection on validity scales. Validity *T* and sometimes raw scores are included because there is extensive research on optimal cutoff scores.

During the interpretive process, do not merely note the meanings of the individual scales but also examine the overall pattern or configuration of the test and note the relative peaks and valleys. Typical configurations, for example, might include the “conversion V,” reflecting a possible Conversion Disorder, or elevated Scales 4 and 9, which reflect a high likelihood of acting-out behavior. Note especially any scales greater than 65 or less than 40 as being particularly important for the overall interpretation. The meaning of 2-point code configurations can be determined by consulting the corresponding section in this chapter (“MMPI-2 2-Point Codes”). When working to understand the meaning of a profile with two or more elevated clinical scales, it is recommended that clinicians read the descriptors for the individual scales as well as relevant 2-point code descriptions. It is also recommended that, when reading about elevations on single scales, clinicians read the meanings of high and low elevations, as well as the more general information on the relevant scale. Further elaboration on the meaning of the scale elevations and code types can be obtained by scoring and interpreting the content scales, Harris-Lingoes and *Si* subscales, supplementary scales, restructured clinical scales, and/or the critical items; these scales are discussed later in this chapter. When interpretive information is available, clinicians can examine an individual’s profile in combination with the requirements of the referral questions to determine relevant descriptions for each of these areas.

Many of the client descriptions focus on client deficits. As a result, clinicians often struggle to translate these interpretations into everyday, client-friendly language. To assist with this, client feedback statements derived from Levak et al. (2011) are included in the individual clinical scale descriptions. The language has been selected to be empathic, to enhance rapport, and to increase the possibility of client growth. These statements can also be edited to develop more client-focused interpretations for

Table 7.2 Minnesota Multiphasic Personality Inventory—2 and Minnesota Multiphasic Personality Inventory—Adolescent Content Scales

Name	Abbreviation	No. of items
MMPI-2 content scales		
Anxiety	ANX	23
Fears	FRS	23
Obsessiveness	OBS	16
Depression	DPS	33
Health concerns	HEA	36
Bizarre mentation	BIZ	23
Anger	ANG	16
Cynicism	CYN	23
Antisocial practices	ASP	22
Type A	TPA	19
Low self-esteem	LSE	24
Social discomfort	SOD	24
Family problems	FAM	25
Work interference	WRK	33
Negative treatment indicators	TRT	26
MMPI-A content scales		
Adolescent-Anxiety	A-anx	21
Adolescent-Obsessiveness	A-obs	15
Adolescent-Depression	A-dep	26
Adolescent-Health	A-hea	37
Adolescent-Alienation	A-aln	20
Adolescent-Bizarre Mentation	A-biz	19
Adolescent-Anger	A-ang	17
Adolescent-Conduct Problems	A-con	23
Adolescent-Cynicism	A-cyn	22
Adolescent-Low Self-Esteem	A-lse	18
Adolescent-Low Aspirations	A-las	16
Adolescent-Social Discomfort	A-sod	24
Adolescent-Family Problems	A-fam	35
Adolescent-School Problems	A-sch	20
Adolescent-Negative Treatment Indicators	A-trt	26

use in actual reports. For example, the statement “Currently, you are very preoccupied with your physical health ...” can become “The client is currently preoccupied with his physical health ...” (see description under “Scale 1. Hypochondriasis (Hs)” later in this chapter). Note that Scales 5 and 0 are not considered clinical scales; as a result, no client feedback statements have been included.

Clearly Defined Profiles

As noted previously, a clearly defined code type is indicated by both a high elevation and either single scales that are elevated with no other “competing” scale elevations (so-called spike profiles) or clear code types in which the elevated scales in the code types similarly do not have competing scales that are close to the degree of elevations of the scales in the code. Well-defined elevations indicate greater validity of the relevant descriptors (McNulty et al., 1998). In addition, they are more likely to be stable over time (high test-retest reliability).

Poorly Defined Profiles

If the elevation is not particularly high (generally $T = 60-65$), the interpretations need to be modified by either toning down the descriptors to a more normal level or deleting the more extreme descriptors. Often the content, Harris-Lingoes, restructured clinical, and supplementary scales can be useful in understanding the meaning of elevations in the $T = 60$ to 64 range. If the profile is poorly defined because there are additional scales that “compete” with the scales in the code type (e.g., 27/72 code type but with Scales 1 and 8 also elevated nearly as high as Scales 2 and 7), several strategies need to be used. The safest and most conservative strategy is to consider descriptors that occur in common among all the different elevated scales as the most valid (e.g., anxiety is likely to be a common descriptor for elevations on Scales 1, 2, 7, and 8; this is strengthened if 7 is the most highly elevated scale). In addition, examiners need to make an effort to understand and integrate the interpretations given under each of the individual scale descriptions. Furthermore, the meanings of alternative code type combinations need to be considered and integrated (e.g., if Scales 2, 7, 1, and 8 are all elevated, these code type descriptors need to be considered: 27/72, 18/81, 87/78, 12/21, 17/71, and 28/82). Finally, with poorly defined elevations, it becomes increasingly important to use the content, Harris-Lingoes, supplementary, and restructured clinical scales to more fully understand and refine the meanings of the clinical scale elevations.

Use of Content Scales

The content scales can be used to supplement, extend, confirm, and refine interpretations derived from the basic validity and standard clinical scales. Furthermore, some of the content scales (e.g., TPA/Type A, WRK/Work Interference) provide additional information not included in the clinical scales. The adult content scales are divided into the clusters of internal symptoms, external aggressive tendencies, negative self-view, and general problem areas. Similarly, the adolescent content scales are divided into scales reflecting interpersonal functioning, treatment recommendations, and academic difficulties (see “MMPI-A Content Scales” section later in this chapter).

Harris-Lingoes and Si Subscales

To understand which personality and clinical variables of a person might have been responsible for elevating the clinical scales, clinicians might wish to selectively use the rationally devised Harris-Lingoes and Social Introversion subscales. These scales (or subscales) organize clusters of content-related items so that the different dimensions

of the scales can be more clearly differentiated. For example, it might be found that an elevation on Scale 4 (Psychopathic Deviate) resulted primarily from family discord. In contrast, criminal acting out might be suggested by subscale elevations on authority conflict and social imperturbability. These findings would then have implications for both interpretations and case management (see “MMPI-2 Harris-Lingoes and *S_i* Subscales” section later in this chapter).

Critical Items

Clinicians may also wish to evaluate the meanings of content related to specific items the client has endorsed by investigating critical items (see the “MMPI-2 Critical Items” section).

Supplementary (Including Psychopathology Personality Five/PSY-5) Scales

The empirically derived supplementary scales can be used, similar to the content and Harris-Lingoes scales, both to refine the meanings of the clinical scales and to add information not included in the clinical scales. A detailed description of these scales is presented in the “Psychopathology Five Scales” section of the MMPI-2-RF interpretation.

Restructured Clinical Scales

Although primarily used in the MMPI-2-RF, the restructured clinical scales represent purer measures of the clinical scales. These purer measures were developed by extracting the common factor of demoralization. As such, elevations on these scales represent clearer measures of the types of variables the basic clinical scales are trying to measure. Detailed descriptions of the Restructured Clinical scales are presented within the “MMPI-2-RF Restructured Clinical Scales” section later in this chapter.

It is recommended that the Restructured Clinical scales be used to refine the meanings of the clinical scales. For example, a client might have elevations on the so-called psychotic or right side of an MMPI-2 clinical scale profile (Scales 6, 8, and 9). These scores may lead a clinician to wonder if the elevations are due primarily to actual psychotic symptoms or if the elevations are merely due to general distress and demoralization. If the corresponding Restructured Clinical scales (see Scales RC6, RC8, and RC9) were quite low in comparison to the clinical scales (6, 8, and 9), a clinician could reasonably infer that the “psychotic” scales were elevated primarily due to demoralization (rather than the presence of actual psychotic symptoms), especially with elevation on RCd.

Low Scale Scores

Low scale scores (below *T* score of 35 or 40) on the clinical scales may represent strengths, and these strengths might correspond in an opposite direction to the interpretations for the high scores. For example, a low score on Scale 1 (Hypochondriasis) might suggest an absence of physical complaints and health-related concerns. However, research in this area is both minimal and equivocal. As a result, this area of interpretation has not been included in the interpretive statements. However,

sufficient research is available for low scores on some of the validity scales and scales 5 (Masculinity/Femininity) and 0 (Social Introversion) such that interpretations have been included.

Specific Interpretive Guidelines Organized around Symptom Domains

The following topic areas and interpretive strategies are intended to be basic, rule-of-thumb approaches to help guide hypothesis generation around specific areas. There are certainly other relevant areas, but the ones listed can generally be considered the most important. While these guidelines will serve to alert clinicians to specific areas, clinicians will still need to investigate these areas in far more depth by consulting relevant scale descriptors and patterns between scales. Clinicians may also wish to consult one of the MMPI-2/MMPI-resources listed in the Recommended Reading section to further extend and expand on the meanings of different profiles.

Suppression (Constriction). Scales 5 (*Mf*) and 0 (*Si*) are sometimes referred to as suppressor scales because, if either or both are elevated, they tend to suppress, or “soften,” the expression of characteristics suggested by other elevated scores.

Acting Out (Impulsivity). In contrast to Scales 5 (*Mf*) and 0 (*Si*), Scales 4 (*Pd*) and 9 (*Ma*) are sometimes referred to as “releaser” or “excitatory scales.” If one or both are elevated, the person is likely to act out difficulties. This hypothesis is further strengthened if 0 (*Si*) is also quite low.

Internalizing Coping Style. Similar to the preceding two guidelines are indicators of internalizing versus externalizing coping styles. If the combined scores for Scales 4 (*Pd*), 6 (*Pa*), and 9 (*Ma*) are lower than the combined scores for 2 (*D*), 7 (*Pt*), and 0 (*Si*), the individual can be considered to have an internalizing coping style.

Externalizing Coping Style. In contrast to the preceding, an individual who has combined scores on 4 (*Pd*), 6 (*Pa*), and 9 (*Ma*) that are greater than his or her combined scores on 2 (*D*), 7 (*Pt*), and 0 (*Si*) can be considered to have an externalizing coping style.

Overcontrol (Repression). Rigid overcontrol of impulses, particularly hostility, is suggested by elevations on 3 (*Hy*) and the *O-H* (Overcontrolled Hostility) supplementary scale.

Anger (Loss of Control). Angry loss of control is suggested by elevations on the ANG (Anger) content scale.

Subjective Distress. A general check on the degree of subjective stress a person is encountering can be determined by noting the degree to which scales 2 (*D*) and 7 (*Pt*) are elevated.

Anxiety. Elevations on Scale 7 (*Pt*), especially if 7 (*Pt*) is greater than 8 (*Sc*), suggest anxiety.

Depression. A high score on 2 (*D*) combined with a low score on 9 (*Ma*) is particularly indicative of depression.

Mania. A high score on 9 (*Ma*) combined with a low score on 2 (*D*) suggests mania.

Psychosis. A high score on 8 (*Sc*) and BIZ (Bizarre Mentation), especially if 8 (*Sc*) is 10 points or more higher than 7 (*Pt*), suggests psychosis.

Confusion and Disorientation. Elevations above $T = 80$ on *F*, 8 (*Sc*), and 7 (*Pt*) suggest a confused, disoriented state. Confusion can also be suggested if the mean for all

eight clinical scales (this excludes Scales 5 and 0, as these are not strictly clinical scales) is greater than $T = 70$.

Suspicion and Mistrust. If 6 (*Pa*) is moderate to highly elevated and especially if it is the highest scale, suspicion and mistrust is strongly indicated.

Introversion. Introversion is indicated by elevations on the 0 (*Si*) scale.

Obsessiveness. Obsessiveness is indicated by elevations on 7 (*Pt*; especially when this is the highest point) and elevations on the OBS (Obsessiveness) content scale.

Cynicism. Cynicism is indicated by elevations on the CYN (Cynicism) content scale.

Drug or Alcohol Problems. Elevations on Scales 4 (*Pd*), 2 (*D*), and 7 (*Pt*) are consistent with (although not diagnostic of) drug- and alcohol-related problems. Lifestyle and personality patterns consistent with and suggesting proneness to drug and alcohol patterns are indicated by elevations on MAC-R and the Alcohol Potential Scale (APS). Clear awareness and open discussion of alcohol and/or drug problems are indicated by elevations on the Alcohol Acknowledgment Scale (AAS).

Quality and Style of Interpersonal Relations. Scales that are most useful for understanding the patterns of interpersonal relations include:

- 0 (*Si*; level of sociability, shyness, social avoidance, alienation).
- Social Discomfort Scale (SOD; social discomfort).
- 1 (*Hs*; complaining, critical, demanding, indirect expression of hostility, passive, preoccupied with self).
- 4 (*Pd*; good first impressions but use others for their own needs, outgoing, talkative, energetic but also shallow and superficial, and impulsive).
- 6 (*Pa*; moralistic, suspicious, hypersensitive, resentful, guarded).
- 8 (*Sc*; isolated from social environment, seclusive, withdrawn, inaccessible, feels misunderstood).
- Marital Distress Scale (MDS; presence of marital conflict).
- Dominance (*Do*; assertive, dominant, takes the initiative, confident).

Step 7. Provide Diagnostic Impressions

Although the original MMPI and the MMPI-2/MMPI-A have not been successful in leading directly to diagnosis, they can often contribute considerable information relevant to diagnostic formulations. In the section on code types, possible *DSM-5* diagnoses consistent with each code type have been included. Clinicians should consider these, along with additional available information, to help make an accurate diagnosis. In some contexts and for some types of referral questions, formal diagnosis will be relevant; but for other contexts and referral questions, formal diagnosis will be neither required nor appropriate (e.g., employment screening). A further review of the considerations and guidelines described in Step 6 might be useful in extracting relevant information for diagnosis.

Step 8. Elaborate on Treatment Implications and Recommendations

Often, one of the most valuable services a practitioner can provide is to predict the client's likelihood of benefiting from interventions. This typically means elaborating

on the person's strengths and weaknesses, level of defensiveness, ability to form a treatment relationship, predicted response to psychotherapy (note especially *Es* [Ego Strength] and TRT scales), antisocial tendencies, and level of insight. Much of this information is summarized at the ends of the subsections on scale elevations and code types. If doing extensive work with specific types of clients, clinicians might need to expand on the knowledge relating to types and outcome of treatments by referring to the extensive research base that is available (e.g., chronic pain, substance abuse, outcomes related to specific code types). Butcher and Perry's (2008) *Personality Assessment in Treatment Planning: Use of the MMPI-2 and BTPI* can be particularly helpful in this regard, as can Levak et al.'s (2011) *Therapeutic Feedback with the MMPI-2: A Positive Psychology Approach*. Treatment responsiveness might be extended into providing suggestions for tailoring specific interventions for client profiles and types of problems. Reviewing the areas, considerations, and guidelines described in Step 6 might be useful in extracting information relevant to treatment planning. A further useful resource in this process is Maruish's (2004) *Use of Psychological Testing for Treatment Planning and Outcome Assessment*. In addition, Levak et al.'s (2011) book and a manual by Finn (1996) offer descriptions of how to use MMPI-2 feedback as a collaborative and therapeutic intervention (see also Finn, Fischer, & Handler, 2012).

MMPI-2 COMPUTERIZED INTERPRETATION

Computerized interpretation systems are an important and frequently used adjunct to MMPI interpretation. The number of such services has grown considerably since 1965, when the first system was developed by the Mayo Clinic. Major providers are Pearson Assessments (previously National Computer Systems), Psychological Assessment Resources, Caldwell Report, Western Psychological Services, Psychometric Software, Psych Screen, Automated Assessment Associates, and Behavior Data. A description and evaluation of many of these services are included in past and current editions of the *Mental Measurements Yearbook* (the most recent/nineteenth edition was edited by Carlson, Geisinger, & Jonson, 2014) and a review by J. E. Williams and Weed (2004). The best sources will be the most recent listings found in test publisher catalogs (see Appendix A) or on service provider websites.

Caution in the use of different computer-based interpretive systems is important because the interpretive services and software packages are highly variable in terms of quality, and most have untested or only partially tested validity. Many do not specify the extent to which they were developed using empirical guidelines versus clinical intuition. Each computerized system has a somewhat different approach. Some provide screening, descriptive summaries, and cautions related to treatment, whereas others provide extensive elaborations on interpretations or may provide optional interpretive printouts for the clients themselves. Even the best programs produce a combination of accurate and inaccurate interpretations (Butcher, Perry, & Hahn, 2004).

The rationale behind computerized systems is that they are efficient and can accumulate and integrate large amounts of information derived from the vast literature on the MMPI, which even experienced clinicians cannot be expected to recall. Additionally, the complex hand-scoring of the MMPI-2 is prone to human error, which

is significantly reduced with computer-based scoring. However, questions have been raised regarding misuse (Groth-Marnat, 1985; Groth-Marnat & Schumaker, 1989). In particular, computerized services are limited to standard interpretations and are not capable of integrating the unique variables usually encountered in dealing with clinical cases. This is a significant factor, which untrained personnel may be more likely either to overlook or to evaluate inadequately. In response to these issues, the American Psychological Association developed a set of guidelines to ensure the proper use of computerized interpretations (American Psychological Association, 1986, 1991). It should be stressed that although computerized systems can offer information from a wide variety of accumulated data, their interpretations are still not end products. Like all test data, they need to be placed in the context of the client's overall background and current situation and integrated within the framework of additional test data (see Lichtenberger, 2006; McMinn, Ellens, & Soref, 1999).

MMPI-2 VALIDITY SCALES

The MMPI was one of the first tests to develop scales to detect whether respondents were answering in such a manner as to invalidate the overall results. This tradition has continued and been expanded in the MMPI-2 and MMPI-A. Meta-analyses of studies on the various validity scales generally indicate that they are able to detect faking effectively. Probably the most effective strategy is the *F* scale's ability to detect overreporting of pathology (R. Baer, Kroll, Rinaldo, & Ballenger, 1999; Bagby, Buis, & Nicholson, 1995; Iverson, Franzen, & Hammond, 1995; G. Meyer & Archer, 2001). The *K* scale, while still useful, is somewhat less effective in detecting underreporting (R. Baer, Wetter, & Berry, 1992; Putzke, Williams, Daniel, & Boll, 1999). However, adding supplementary validity scales (Social Desirability scale, Superlative scale) to *L* and *K* can serve to increase the detection of underreporting (Bagby, Rogers, Nicholson, et al., 1997). Despite the near consensus related to the accuracy of detection, a concern is that a wide range of cutoff scores are recommended depending on the group being assessed (Bagby et al., 1994, 1995; L. Stein, Graham, & Williams, 1995). For example, optimal cutoff scores for normals faking bad are lower than psychiatric patients faking bad (Berry, Baer, & Harris, 1991; J. R. Graham et al., 1991). An unresolved issue is whether normals who are motivated to fake bad and are given information on how to fake (e.g., symptom patterns of individuals with posttraumatic stress disorder, paranoid schizophrenia) can avoid detection. Some research indicates that, even with motivation and a clear strategy, they still cannot avoid detection (Wetter, Baer, Berry, Robinson, & Sumpter, 1993), whereas other research suggests that strategic (informed) fakers can consistently produce profiles that are indistinguishable from those of true patients (R. Rogers, Bagby, & Chakraborty, 1993; Wetter & Deitsch, 1996). Attempts to fake bad might be particularly likely to succeed if subjects are given information on the design and intent of the validity scales (Lamb, Berry, Wetter, & Baer, 1994) and are familiar with the type of disorder they are faking (Bagby, Rogers, Buis, et al., 1994).

It should be noted that the MMPI-2 provides the option of profile sheets that include *K* corrections or sheets that omit this procedure. The MMPI-A does not include the *K*

correction on its profile sheets because, in some contexts, particularly those for adolescents, the *K* correction is not appropriate (Colby, 1989).

? “Scale” (Cannot Say, *Cs*)

The ? scale (abbreviated by either ? or *Cs*) is not actually a formal scale but merely represents the number of items left unanswered on the profile sheet. The MMPI-2 does not include a column for profiling a ? (*Cs*) scale but merely provides a section to include the total number of unanswered questions. The usefulness of noting the total number of unanswered questions is to provide one of several indices of a protocol’s validity. If 30 or more items are left unanswered, the protocol is most likely invalid, and no further interpretations should be attempted. This is simply because an insufficient number of items have been responded to, which means less information is available for scoring the scales. Thus, less confidence can be placed in the results. To minimize the number of “cannot say” responses, the client should be encouraged to answer all questions.

High Number of ? (30+)

- Difficulties with reading, psychomotor retardation, indecision, confusion, or extreme defensiveness (consistent with severe depression, obsessional states, extreme intellectualization, or unusual interpretations of the items).
- Legalistic overcautiousness or a paranoid condition.
- Perception that the unanswered items are irrelevant.

VRIN (Variable Response Inconsistency Scale)

The VRIN includes pairs of selected questions that would be expected to be answered in a consistent manner if the person is approaching the testing in a valid manner. Each pair of items is either similar or opposite in content. It would be expected that similar items would be answered in the same direction. If a person answers in the opposite direction, then it indicates an inconsistent response and is, therefore, scored as 1 raw score point on the VRIN scale. Pairs of items with opposite contents would be expected to be answered in opposite directions. If, instead, these pairs are answered in the same direction, this would represent inconsistent responding, which would also be scored as 1 raw score point on the VRIN scale.

High VRIN (MMPI-2 *T* > 79; MMPI-A *T* > 74) or Moderate (MMPI-2 *T* 70–79; MMPI-A *T* 70–74)

- Indiscriminate responding; profile should be considered invalid and should not be interpreted (especially if *F* is also high).

TRIN (True Response Inconsistency Scale)

The MMPI-2 and MMPI-A TRIN scale is like the VRIN scale in comprising pairs of items. However, only pairs with opposite contents are included. This means there would be two ways for a person to obtain a response that would be scored on the TRIN scale. A “True” response to both items would indicate inconsistency and would

therefore be scored as plus 1 raw score point. A “False” response to both pairs would also indicate inconsistency but would be scored as minus 1 raw score point (negative scores are avoided by adding a constant).

Very High (MMPI-2 $T > 79$; MMPI-A $T > 74$) or Moderate (MMPI-2 T 70–79; MMPI-A T 70–74)

- Person is indiscriminately answering “True” to the items (acquiescence or yea-saying).

***F* Scale (Infrequency)**

The *F* (Infrequency) scale measures the extent to which a person answers in an atypical and deviant manner. The MMPI and MMPI-2 *F* scale items were selected based on their endorsement by less than 10% of the population. Thus, from a statistical definition, they reflect nonconventional thinking. This nonconventional thinking may include endorsing items that all rules should be thrown away or that the examinee would like to visit novel places. These items do not cohere around any particular trait or syndrome. High scores indicate the examinee is answering in a scorable direction to a wide variety of unusual characteristics. As might be expected, high scores on *F* are typically accompanied by high scores on many of the clinical scales. High scores can often be used as a general indicator of pathology. In particular, high scores can reflect unusual feelings caused by some specific life circumstance to which the person is reacting. This might include grieving, job loss, or divorce. A person scoring high may also be “faking bad,” which could serve to invalidate the protocol. No exact cut-off score is available to determine whether a profile is invalid or is accurately reflecting pathology. Even *T* scores from 70 to 90 do not necessarily reflect an invalid profile, particularly among prison or inpatient populations. In general, moderate elevations represent an openness to unusual experiences and possible psychopathology, but it is not until more extreme elevations that an invalid profile is suspected. Further information can be obtained by consulting the *F* back scale (see the section titled “*Fb* (*F* back) Scale [MMPI-2]; *F1* and *F2* [MMPI-A]”).

The 66-item MMPI-A *F* scale was constructed similar to the MMPI-2 *F* scale. However, because adolescents are more likely to endorse unusual experiences, a more liberal criterion of 20% endorsement was used for inclusion. The MMPI-A *F* scale was further divided into *F1* scales to assess validity for the first portion of the booklet (clinical scales) and *F2* to assess the last portion of the book (supplementary and content scales; *F1* and *F2*).

High Scores on *F* (approximately $T > 99$; fake bad cutoff for inpatients = 100, cutoff for outpatients $T = 90$, cutoff for nonclinical settings $T = 80$; cutoff for MMPI-A $T = 79$)

- Invalid profile, possibly caused by clerical errors in scoring, random responding, false claims by the client regarding symptoms, resistance to testing, malingering.
- *Extremely high F (100+)*. May possibly accurately reflect psychopathology, but this will correspond with possible hallucinations, delusions of reference, poor judgment, disorientation, restlessness, dissatisfaction, and/or extreme withdrawal (check for consistency with history).

Moderate Scores ($T = 70-90$)

- Attempt to draw attention to distress as a cry for help (and are in need of assistance).
- Unconventional and unusual thoughts, may be rebellious, antisocial, and/or having difficulties in establishing a clear identity.
- *Slightly elevated ($T = 65$ to 75) and person does not seem to be pathological.* He or she might be curious, complex, psychologically sophisticated, opinionated, unstable, and/or moody.

Low Scores on F

- Clients perceive the world as most other people do.
- Possible denial of difficulties if their history indicates psychopathology (“faking good”; note the relative elevation on K and L).

 Fb (F back) Scale (MMPI-2); $F1$ and $F2$ (MMPI-A)

The 40-item MMPI-2 Fb was designed to identify a “fake bad” mode of responding for the last 197 items. This might be important because the traditional F scale was derived only from items taken from what are now the first 370 questions on the MMPI-2. Without the Fb scale, no check on the validity of the later questions would be available. It might be possible for a person to answer the earlier items accurately and later change to an invalid mode of responding, especially because the length of the test can lead to fatigue partway through. This is important for the supplementary and content scales because many of them are derived either partially or fully from the last 197 questions. The Fb scale was developed in the same manner as the earlier F scale in that items with low endorsement frequency (less than 10% of nonpatient adults) were included. Thus, a high score suggests the person was answering the items in an unusual mode. As with the F scale, this could indicate either generalized pathology or that the person was attempting to exaggerate his or her level of symptomatology.

Somewhat similar to the MMPI-2, the MMPI-A includes a 66-item F scale that is divided into $F1$ and $F2$ subscales. The $F1$ scale is composed of 33 items, all of which appear on the first half (initial 236 items) of the MMPI-A booklet, and relates to the standard clinical scales. In contrast, the 33-item $F2$ scale is composed of items on the last half of the booklet (final 114 items) and relates to the supplementary and clinical scales. The $F1$ and $F2$ scales can be interpreted in much the same way as for F and Fb on the MMPI-2. However, the Fb scale has not been found to be as effective a predictor of malingering as the F scale (Iverson et al., 1995).

High Fb (and $F1$ and $F2$; $T > 89$ for nonclinical settings, 109 for clinical settings)

- Possible exaggeration of psychopathology (see considerations under “ F Scale [Infrequency]” above).

 Fp (Infrequency-Psychopathology) Scale

Because the F scale is typically elevated among psychiatric patients, it is often difficult to differentiate between persons with true psychopathology and those who have some psychopathology but are nonetheless faking bad. This is particularly true if the

psychopathology is quite severe. The history of the person (e.g., degree of preexisting psychopathology) and context of the referral (e.g., possible gain for faking bad) can often be quite useful in making this distinction. To further assist with this differentiation, Arbisi and Ben-Porath (1995) developed a set of 27 items that were answered infrequently even by psychiatric inpatients. (In contrast, the *F* scale was developed from infrequently answered questions by the normative sample.)

High *F_p* ($T > 93$ for men, $T > 96$ for women)

- High probability of faking or exaggerating psychopathology, even among psychiatric patients.

Fake Bad Scale (*FBS*)

The Fake Bad Scale (*FBS*) was developed in the hopes that it could detect personal injury claimants who were exaggerating their difficulties (Lees-Haley, English, & Glenn, 1991). Research has been equivocal with concerns related to false positives (R. Rogers, Sewell, Martin, & Vitacco, 1999). In contrast, other studies have provided more supportive results (Greiffenstein, Fox, & Lees-Haley, 2007) and indicate it is one of the best MMPI-2 scales for detecting faking (N. W. Nelson, Sweet, & Demakis, 2006).

High *FBS* (moderately indicative if raw score > 21 ; more strongly indicated by a raw score > 27)

- *Fake bad/malingering*. Raw scores of 28 or higher reduce the possibility of false positives.

***L* (Lie) Scale**

The *L* or Lie Scale consists of 15 items that indicate the extent to which a client is attempting to describe him- or herself in an unrealistically positive manner. Thus, high scorers describe themselves in an overly idealized manner. The items consist of descriptions of relatively minor flaws to which most people are willing to admit. Thus, persons scoring high on the *L* scale might state that they never get angry or that they like everyone they meet.

High Scores on *L* ($T > 64$)

- Person is describing self in an overly favorable light due to conscious deception.
- Person is describing self in an overly favorable light due to an unrealistic view of him- or herself; may be inflexible, unoriginal, and unaware of the impressions he or she makes on others; perceives the world in a rigid, self-centered manner.
- Poor insight due to denial of flaws.
- Low tolerance of stress.
- Poor candidates for psychotherapy.
- Extremely high scores would suggest that such persons are ruminative, extremely rigid, and will experience difficulties in relationships (e.g., individuals with paranoia who place considerable emphasis on denying their personal flaws and instead project them onto others).

- Extremely high scores might be due to conscious deception by antisocial personalities.

Low Scores on *L* ($T = 35-45$)

- Frank and open regarding responses to items.
- Able to admit minor faults in themselves, may also be articulate, relaxed, socially ascendant, and self-reliant.
- Possibly somewhat sarcastic and cynical.

***K* (Correction) Scale**

The *K* scale was designed to detect clients who are describing themselves in overly positive terms. It therefore has a similarity with the *L* scale. The *K* scale, however, is more subtle and effective. Whereas only naive, moralistic, and unsophisticated individuals score high on *L*, more intelligent and psychologically sophisticated persons might have somewhat high *K* scores and yet be unlikely to have any significant elevation on *L*.

Moderate scorers often have good ego strength, effective emotional defenses, good contact with reality, and excellent coping skills. Typically, they are concerned with, and often skilled in, making socially acceptable responses. As might be expected, *K* scores are inversely related to scores on Scales 8, 7, and 0. Elevations on *K* can also represent ego defensiveness and guardedness. This might occur with persons who avoid revealing themselves because of their personality style or because something might be gained by conveying a highly favorable impression (e.g., employment, child custody evaluations). There is no clear cutoff for differentiating among positive ego strength (adjustment), ego defensiveness, and faking good. A general guideline is that the more ego-defensive the person is, the more likely it is that some of the clinical scales might also be elevated. Helpful information can be obtained through relevant history and the context of the testing (e.g., legal proceedings, employment evaluation).

Because a defensive test-taking approach is likely to suppress the clinical scales, a *K* correction is added to five of the MMPI-2 clinical scales (1/*Hs*, 4/*Pd*, 7/*Pt*, 8/*Sc*, 9/*Ma*) to compensate for this defensiveness. This correction is obtained by taking a designated fraction of *K* and adding it to the relevant scale. However, the basis of the *K* correction has been called into question. It has been omitted from the MMPI-A, and the MMPI-2 contains separate scoring sheets with and without the *K* correction so that examiners can decide whether they wish to use it.

High Scores on *K* ($T > 65$ or 70)

- Person is attempting to describe self in an overly favorable light, denying difficulties.
- May have answered false to all items (naysaying; check TRIN and VRIN).
- If profile is considered valid, person is presenting an image of being in control and functioning effectively, but he or she will overlook any faults he or she might have.
- Likely to have poor insight and resist psychological evaluation; limited benefit from psychotherapy.
- Intolerant of nonconformity in others, may perceive nonconformists as weak.

- Use of denial, poor insight, unaware of the impression he or she makes on others.
- Shy, inhibited, low level of social interaction (check *Si*).

Moderate Scores on *K* ($T = 56-64$)

- Moderate levels of defensiveness.
- Potential positive qualities: independent, self-reliant, express an appropriate level of self-disclosure, have good ego strength, good verbal ability and social skills.
- Might admit to some “socially acceptable” difficulties but minimize other important conflicts.
- Unlikely to seek help.

Low Scores on *K*

- Fake bad profile, exaggeration of pathology (check *F*, *Fb*, *F1*, and *F2*).
- In an otherwise valid profile, client might be disoriented and confused, extremely self-critical, cynical, skeptical, and dissatisfied and have inadequate defenses.
- Poor self-concept, low level of insight.

***S* (Superlative) Scale**

Because the *K* and *L* scales have been found to be only moderately effective in differentiating persons who fake good, the *S* scale was developed in the hopes that it might more accurately identify those persons attempting to appear overly virtuous (Butcher & Han, 1995). The 50 items of the scale were developed by noting the differences in item endorsement between persons in an employment situation who were likely to be presenting themselves in an extremely favorable light (i.e., airline pilots applying for a job) and the responses of the normative sample. The resulting 50 items relate to contentment with life, serenity, affirming human goodness, denial of irritability/anger, patience, and denial of moral flaws. Thus, persons endorsing a high number of these items are presenting themselves as getting along very easily with others, being free from psychological problems, and having a strong belief in human goodness.

The scale does seem to be effective in discriminating nonpatients who were requested to present themselves in an extremely favorable light (pretending they were applying for a highly desired job) from those who were requested to respond in an honest manner (R. Baer, Wetter, Nichols, et al., 1995). However, the *L* and *K* scales are equally as good in detecting clinical populations who are underreporting psychopathology (R. Baer & Miller, 2002; R. Baer, Wetter, & Berry, 1995).

MMPI-2 CLINICAL SCALES

Scale 1. Hypochondriasis (*Hs*)

Scale 1 was originally designed to distinguish those with hypochondriasis from other types of psychiatric patients. Although it can suggest a diagnosis of hypochondriasis, it is more useful as a scale to indicate a variety of personality characteristics that are often consistent with, but not necessarily diagnostic of, hypochondriasis.

High Scores on Scale 1

- Person has high concern with illness and disease.
- Rarely acts out directly but express hostility indirectly, likely to be critical of others.
- Complaints usually related to a wide variety of physical difficulties; complaints are vague and diffuse and often shift to various locations on the body.
- Complaints typically include gastrointestinal complaints, pain, fatigue, and headaches.
- Complaints are used to manipulate and control others, thereby creating interpersonal distress.
- Symptoms are usually not reactions to situational stress but more of long-standing duration.
- Moderate scores may have a true organic basis for the difficulties, but client still is likely to exaggerate physical difficulties.
- Typically experience little overt anxiety.
- Stubborn, pessimistic, narcissistic, and egocentric.
- Immature, pessimistic, sour, whiny, and passive-aggressive.
- Perceived by others as dull, unenthusiastic, ineffective, and unambitious.
- Level of efficiency is reduced, but client is rarely completely incapacitated.
- Overuses the medical system. Histories usually reveal numerous visits to a wide variety of practitioners. Client will recite a long series of symptom complaints (sometimes referred to as an “organ recital”).
- Refuses to believe assurances that difficulties have no organic basis.
- Clinicians should investigate for possible prescription medication abuse.
- Extremely high scores suggest the person has a wide variety of symptom-related complaints, extremely emotionally constricted, possibly consistent with psychotic-like features (schizoid, schizoaffective, schizophrenia, psychotic depression) with bodily delusions—check elevations on Scales 6, 7, 8, and 9.

Related Scale Elevations

- Often elevated along with Scales 2, 3, and 7; reflect corresponding levels of depression, denial, conversions, or anxiety states.
- *Scale 7 is elevated.* Indicates better prognosis for psychotherapy. Client’s level of anxiety is high enough to motivate him or her to change.
- “*Conversion V*” (elevations on Scales 1 and 3 with a significant lowering of 10 or more points on 2). Person converts psychological conflicts into bodily complaints (see 13/31 code type).

Treatment Implications

- Has often rejected and criticized the “help” that has been offered to him or her.
- Resists any suggestion that his or her difficulties are even partly psychologically based.
- Psychotherapy is usually difficult due to client’s poor insight.

- Pessimistic about being helped, argumentative with professional staff (confirm/disconfirm this by checking the TRT/Negative Treatment Indicators scale).
- Client requires repeated assurance that he or she has been well understood and that symptoms will not be ignored.
- Framing interventions with biomedical terminology may make interventions more acceptable (e.g., biofeedback procedures might be described as “neurological retraining”).

Client Feedback Statements[†]

Your profile suggests that you are feeling a number of physical symptoms that may be frightening you. You may experience pains, weaknesses, insomnia, fatigue, tremors, and stomach upsets. Whenever you are stressed, these symptoms may become more severe. Physical symptoms such as nausea, headaches, and dizziness can come and go, sometimes taking you by surprise . . . perhaps you’re worrying that your physical problems reflect a serious medical problem that could lead to disability and even death. You may experience symptoms of depression such as anxiety, difficulties with concentration and memory, and a loss of interest in sex. Your sleep may be disturbed, and you may experience rapid changes in weight. You may have become inefficient, unable to get things accomplished the way you would like to. It may be hard for you to enjoy much right now, and even when things are going well you may find yourself feeling a dull sense of unhappiness. At other times, you may feel defeated and quite down . . . You may find yourself worried that some physical sensation is a sign that there is something really wrong with you. Much of the time you feel a sense of stress, so that it’s hard to relax, to switch off your mind, and to be in the moment . . . People become rigid about doing things a certain way if they’ve been afraid of physical infirmity. You may have a tendency to get quite stubborn about doing things a certain way, especially if you feel that doing them that way helps you feel safer or less physically ill. Others may see your demands as somewhat rigid and inflexible.

MMPI-A Considerations

The preceding descriptors are relevant for adolescent profiles. They also suggest school-related difficulties. Girls are likely to experience family problems (e.g., marital disagreements, financial concerns) and eating disorders. However, elevations on this scale are relatively rare among adolescents.

Scale 2. Depression (*D*)

The 57 items of Scale 2 relate to brooding, physical slowness, subjective feelings of depression, mental apathy, and physical malfunctioning. High scores indicate difficulties in one or more of these areas. Patients seeking inpatient psychiatric treatment are most likely to have Scale 2 as the highest point on their profiles. As would be expected, elevations on 2 typically decrease after successful psychotherapy. The relative elevation on Scale 2 is the single best predictor of a person’s level of satisfaction with life, sense of security, and degree of comfort. Persons who score high on 2 are usually described as self-critical, withdrawn, aloof, silent, and retiring. Adolescents typically score slightly lower than nonpatient adults, whereas older adults score 5 to 10 points higher.

High Scores on Scale 2

- Suggests acute depression (particularly if 2 is the only high point).
- Confronting difficulties with pessimism, helplessness, and hopelessness; this may be part of characteristic personality features exaggerated due to current problems.
- Sense of inadequacy, poor morale, difficulty concentrating; may be severe enough to create difficulties in working effectively.
- Depression can be seen as both a symptom and a means of coping by numbing self to future painful feelings or situations.
- Retiring, shy, aloof, timid, and inhibited, but also irritable, high-strung, and impatient.
- Highly sensitive to criticism.
- Avoids confrontations at all costs, possibly avoiding interpersonal relationships in general.
- Conventional, cautious, passive, and unassertive; higher scores indicate an exaggeration of these trends.
- Excessive worry over even minor problems.
- Possibly impaired ability to deal effectively with interpersonal problems.
- Possible psychomotor retardation, lethargy, and withdrawal.
- Possible preoccupation with death and suicide; check to determine possible inpatient treatment based on whether client is danger to self.

Related Score Elevations

- *Elevated Scales 1, 2, and 3 are referred to as the neurotic triad.* Wide variety of complaints, including not only depression but also somatic complaints, irritability, difficulties with interpersonal relationships, work-related problems, and general dissatisfaction (see code types 12/21, 13/31, 23/32).
- *Associated elevation on 7 (2 and 7 referred to as the distress scales).* Index of personal pain, anxiety, and discomfort; tense, nervous, introjective, self-critical; favorable sign for psychotherapy since person is motivated to change and is introspective and self-aware (see code type 27/72).
- *Elevations on Scales 2 and 8.* Depression is characterized by unusual thoughts, disaffiliation, isolation, and alienation (see code type 28/82).
- *Corresponding elevations on 1 (also HEA, and Harris-Lingoes D3/Physical Malfunctioning).* Variety of somatic complaints, including feeling sluggish, tense, low energy.

Treatment Implications

- Check for whether external (reactive) or internal (endogenous) factors are responsible for depression.
- Check for relative contribution of cognitions, social support, and the prevalence of vegetative symptoms; focus treatment accordingly.
- Check for suicide potential, particularly if the elevations are high to extremely high with corresponding elevations on 4, 7, 8, and/or 9 (note that no clear “suicidal profile” accurately predicts suicide). Any suggestion of suicidal tendencies on

the profile should be investigated further through a careful assessment of additional relevant variables (demographics, presence, clarity, lethality of plan, etc.).

- Moderate depression possible positive sign for psychotherapy since client is likely to be highly motivated (but check possible negative indicators with elevations on TRT/Negative Treatment Indicators, *L*, *K*, and 1), but extremely high score may indicate client is too depressed to experience sufficient motivation to change.

Client Feedback Statements[†]

Your profile shows that you are currently feeling quite down and depressed. Sometimes people with depression become used to it and no longer realize how depressed they are; in other cases, people experience great discomfort. Often when people get depressed, they also become anxious. They experience this anxiety as a constant sense of dread, as if something bad is about to happen When people are depressed, they often feel a lack of energy and low motivation. Things that, in the past, seemed to take little energy may now seem overwhelming. You may have to push yourself to engage in even the simplest activities that others might find pleasurable. You may find yourself dreading doing even the smallest chores Depression is associated with difficulties with concentration, memory, and general alertness and attention. You may reread the same thing without comprehending it, and you may be unable to remember what you did earlier in the day or the day before. You may even become fearful that you are somehow losing your mind. Generally, these symptoms decrease once the depression is treated Depression is often associated with a loss of hope. You may give up hopes and dreams for the future, feeling that it is useless to have desires because you are likely to be disappointed. People may see you as pessimistic, but this reflects your fear that your life is over and your feeling that the future is bleak People with your profile tend to experience a great deal of guilt. Perhaps some recent setback or past losses have left you with feelings of self-blame, feeling that you have ruined your life and that your failures are unforgiveable. Guilt is a painful companion as you remind yourself of your failures. Even if you do something well or if people say positive things toward you, you may feel guilty as if you do not deserve compliments.

MMPI-A Considerations

The preceding MMPI-2 descriptors and use of the Harris-Lingoes scales are also relevant for adolescents, particularly for girls. In addition, high adolescent scores on 2 suggest school-related difficulties (check A-sch/School Problems content scale) and a worsening of arguments with their parents (check A-fam/Family Problems content scale). They are less likely to act out but more likely to report eating problems (especially girls), somatic complaints, and low self-esteem. Interpersonally, they will be introverted with a few friends.

Scale 3. Hysteria (*Hy*)

Scale 3 was originally designed to identify patients who had developed a psychogenically based sensory or motor disorder. The 60 items primarily involve specific physical complaints and a defensive denial of emotional or interpersonal difficulties. The types of physical complaints are generally quite specific and include areas such as fitful sleep, nausea, vomiting, headaches, and heart or chest pains (check HEA/Health Concerns

scale). The important feature of persons who score high on this scale is that they simultaneously report specific physical complaints and use a style of denial in which they may even express an exaggerated degree of optimism. One of the important and primary ways in which they deal with anxiety and conflict is to channel or convert these difficulties onto the body. Thus, their physical complaints serve as an indirect expression of conflicts. Their traits might be consistent with a histrionic personality in that they will demand affection and social support but do so in an indirect and manipulative manner. They are also likely to be socially uninhibited and highly visible. They can easily initiate relationships, yet their relationships are likely to be superficial. They will approach others in a self-centered and naive manner. They might act out sexually or aggressively but have a convenient lack of insight into either their underlying motives or their impact on others. However, Scale 3 is quite heterogeneous in its item composition. The Harris-Lingoes item analysis has divided these Scale 3 items into denial of social anxiety, need for affection, lassitude-malaise, somatic complaints, and inhibition of aggression. If Scale 3 is clearly elevated and a clinician is unclear regarding the meaning of the elevation, it can often be useful to formally score the Harris-Lingoes subscales (see the section on “MMPI-2 Harris-Lingoes and *Si* Subscales” later in this chapter).

High Scores on Scale 3

- Extraverted, dramatic, attention-seeking.
- Highly conforming, immature, naive, childish self-centered, impulsive.
- Strong needs for approval, support, and affection; will attempt to obtain these through indirect and manipulative means; interpersonally indirect.
- Difficulty expressing hostility and resentment.
- Will communicate with others to create an impact rather than to convey specific information.
- Will perceive events globally rather than attend to specific and often relevant details of a situation.
- Low levels of anxiety, tension, depression; rarely report serious psychopathology, such as hallucinations, delusions, and suspiciousness.
- Presence of functionally related somatic complaints.
- Physical difficulties typically worsen in response to increases in stress levels, typically disappear when stress is reduced (especially with $T > 79$).
- Complaints can be either quite vague or quite specific and are of unknown origin.
- Clients will explain symptoms in purely medical terms, will seek medical rather than psychological treatment.
- Use of denial combined with dissociation.
- Low insight; deny difficulties and have a strong need to see themselves in a favorable light.
- Increasingly higher scores reflect exaggeration of denial, somatization, dissociation, immaturity, suggestibility, and low levels of insight.
- Persons with moderate scores may have good levels of adjustment, especially if educated and from higher socioeconomic groups.

- Persons with moderate scores may be presenting a favorable impression for employment (reflects endorsement of items denying any abnormality).

Related Scale Elevations

- Note “Conversion V” (see description in “Scale 1. Hypochondriasis (*Hs*)” section earlier in this chapter; see also code types 12/21, 13/31, and 23/32).
- *High 2 with K*. Person is likely to be inhibited, affiliative, overconventional, have an exaggerated need to be liked and approved of by others (especially if scales *F* and 8 are low).
- *High 3* reduces the likelihood the person will be psychotic, even though Scales 6 and 8 might be relatively high.

Treatment Implications

- Enthusiastic, optimistic initial response to therapy, partially based on strong needs to be liked.
- Slow to gain insight into underlying motives for behavior due to extensive denial and repression, will deny presence of psychological problems.
- Often look for simplistic, medical, concrete, naive solutions.
- Will try to manipulate therapist into supportive, nonconfrontational role; if defenses are challenged, they might become more manipulative, perhaps resorting to complaints of mistreatment and not being understood, possibly even becoming verbally aggressive.
- Core conflicts centered on issues of dependence versus independence.
- Direct suggestion focusing on short-term goals is often effective in creating change.

Client Feedback Statements[†]

Your profile suggests that you may be experiencing a number of health concerns and problems. Perhaps you have occasional headaches, stomachaches, or low back pain. You may experience various pains and weaknesses, dizziness, nausea, fatigue, and other vague and shifting physical symptoms. These symptoms may frighten you and cause you discomfort, but you try to stay brave and positive. Some of these symptoms may be very unnerving, especially if doctors are unable to diagnose what exactly is wrong. What might be particularly confusing is that these symptoms may shift and change, with no one symptom dominating for very long. These physical symptoms may become more severe during times of stress and then suddenly diminish Although you try to stay positive and optimistic and you do a good job of playing the right role, underneath you may feel overwhelmed and anxious, especially when your physical symptoms are worse. It may be hard for you to do things for yourself, and you may feel a need to obtain other people's support and help People with your profile try to be positive and brave, even in the face of pain and discomfort. It is important for you to be seen by people as a cheerful and nice person, and you work hard to avoid conflict. You try to see the best in people, so that sometimes people can disappoint you because you have overlooked or denied their negative attributes. It's important for you to think positively of people, and it's important to you that people like you and see you as a good person. You work hard to get their approval. You try to see

the best in others, turning a blind eye to their failings You tend to repress and deny some of your negative emotions because it's so important for you to be positive, happy, and cheerful and not to upset the people around you. It's as if you have learned to not feel negative emotions to stay connected and close to people.

MMPI-A Considerations

Although the interpretations for adults with elevated Scale 3 can also be made with adolescents, they should be done with caution because of concerns related to questionable validity with this population. In particular, the Harris-Lingoes subscales can help to clarify the meanings of scale elevations. Females (but not males) are still likely to have somatic complaints in response to stress. Males are more likely to have both school problems (check A-sch/School Problems content scale) and a history of suicidal ideation and gestures. However, Scale 3 is rarely a high point among adolescent males.

Scale 4. Psychopathic Deviate (*Pd*)

The purpose of Scale 4 is to assess aspects of the person's general level of social adjustment. The questions deal with areas such as degree of alienation from family, social imperviousness, difficulties with school and authority figures, and alienation from self and society (see ANG/Anger and FAM/Family Problems content scales). The original purpose of the scale was to distinguish those persons who had continuing legal difficulties yet were of normal intelligence and did not report having experienced cultural deprivation. They were people who seemed unconcerned about the social consequences of their behavior and yet did not appear to suffer from neurotic or psychotic difficulties. An important rationale for developing the scale is that high scorers might not be engaged in acting out at the time of testing. In fact, they may often make an initial good impression, which could sometimes be described as charming. Recent friends and acquaintances may not believe that they could even be capable of antisocial behavior. However, under stress or when confronted with a situation that demands consistent, responsible behavior, they would be expected to act out in antisocial ways. Even though they might get caught, these persons would have a difficult time learning from their mistakes.

Different relatively normal groups often have somewhat elevated Scale 4 profiles. This might include counterculture groups, which reflect their relative disregard for the values and beliefs of mainstream culture. Similarly, African Americans often score higher, which might reflect their feelings that many of the rules and laws of the dominant culture are unfair and serve to disadvantage them. Normal persons who are graduate students in the humanities and social sciences often have somewhat elevated scores. More positive characteristics to be found with moderate elevations include frankness, deliberateness, assertion, sociability, and individualism. In addition, normal persons who are extraverted, risk takers, and have unconventional lifestyles (e.g., skydivers, police officers, actors) are also likely to have somewhat elevated Scale 4 profiles.

High Scores on Scale 4

- Problems with persons in authority.
- Frequent relationship and work difficulties.

- Poor tolerance for boredom.
- Angry disidentification with family or society or both.
- Relationships shallow and characterized by recurrent turmoil, difficulty forming long-term loyalties.
- May make an initial good impression; eventually they will have an outbreak of irresponsible, untrustworthy, and antisocial behavior.
- Slow to learn from the consequences of their behavior even when caught.
- When confronted with the consequences of their actions, may experience genuine remorse, but this is usually short-lived.
- Difficulty in learning from experience makes benefiting from psychotherapy difficult.
- Will blame others, particularly their families, when things go wrong.
- High scorers on 4 are often perceived as angry, alienated, impulsive, and rebellious (see ASP/Antisocial Practices content scale) but also outgoing, extraverted, talkative, active, and self-centered.
- Frequent history of involvement with the legal system; extensive alcohol or drug abuse.
- Extremely high scores: aggressive or even assaultive, unstable, irresponsible, self-centered, legal difficulties.
- Moderate scores: adventurous, pleasure-seeking, sociable, self-confident, assertive, unreliable, resentful, and imaginative.

Related Scale Elevations

- *Elevated 4 and 9* (see 49/94 code type and ASP/Antisocial Practices and ANG/Anger content scales). Presence of energy to act on underlying feelings of anger and impulsiveness, history of extensive impulsive behavior, behaviors have resulted in damage to family's reputation, possible involvement with criminal activity. Moderate elevations on Scales 4 and 9 suggest assertive behaviors with possibly a good level of adjustment.
- *High 4 and 8* (see 48/84 code type). Psychotic expression of antisocial behavior.
- *High 4 accompanied by a high 3* (see 34/43 code type). Antisocial behavior might be expressed in covert or disguised methods, person might manipulate others to act out for him or her.
- *High 4 and 2* (see 24/42 code type). Person has been caught performing antisocial acts, feels temporary guilt and remorse for his or her behaviors.

Treatment Implications

- May initially be perceived as good candidates for psychotherapy (they are usually verbally fluent, energetic), but underlying hostility, impulsiveness, and feelings of alienation eventually surface; likely to blame others for the problems they have encountered.
- Initial agreement to psychotherapy may be to avoid negative consequences. When these are removed (e.g., end of parole, spouse moves back in), they usually terminate treatment (see TRT/Negative Treatment Indicators scale).

- Difficulty committing themselves to any, including the therapeutic, relationship.
- If low subjective distress (low Scales 2 and 7), motivation for change is likely to be particularly low.
- Short-term goals that focus on documenting clear behavior change (rather than merely verbalizing it) are beneficial.
- External motivation for therapy (e.g., condition of parole or continued employment) will increase the likelihood that they will follow through on treatment.

Client Feedback Statements[†]

You are willing to challenge the established way of doing things, and you're not afraid to look at things from a new and different perspective. You are an independent, excitement-seeking, and somewhat risk-taking individual who learned, from an early age, to be a survivor You may also have learned that being manipulative is how one gets along in the world. You may see the world as a "dog-eat-dog" place where being "top dog" is the only solution. You may find yourself seeking positions of power and control to avoid others having control over you. It's hard for you to trust other people or to let down your guard and ask others for emotional support. Your fear is that if you reveal your weaknesses to others, they will somehow exploit you by using that against you. Others may see you as more manipulative and devious than you see yourself. As you had to learn from an early age to manipulate your parents to get your needs met, you may have learned to "selectively report," to tell white lies, and even to openly lie as a way of avoiding conflict or negative consequences You probably learned to numb your emotions and not let yourself feel: as a result, you may experience the world as somewhat boring and lacking in intensity and excitement Although people may find you attractive and you enjoy socializing, you find it difficult to allow yourself to be committed, let your guard down, and be emotionally close Sometimes people with your profile have trouble with the law or with authority figures For you to obey the rules, authority figures have to gain your respect. You tend to look for their flaws and weakness, perhaps justifying why you won't conform and obey basic regulations.

MMPI-A Considerations

Adolescents frequently have elevations on Scale 4, and it is often their highest overall scale. A full one-third of the clinical sample used in the development of the MMPI-A had elevations of 65 or more. These generally high scores most likely reflect these individuals' often turbulent attempts to form a sense of identity and achieve independence from their parents. Thus, the elevation might be part of a temporary phase of development rather than a permanent, enduring trait. However, high or extremely high scores still reflect significant levels of pathology. Such scores are associated with delinquents who commit antisocial acts (see A-ang/Anger and A-con/Conduct Problems scales), are in conflict with their families (see A-fam/Family Problems), have school-related difficulties (see A-Sch/School Problems), and are involved with drugs and/or alcohol (see MAC-R, ACK/Alcohol Drug Acknowledgment, and PRO/Alcohol Proneness supplementary scales). Often these individuals report little guilt for this acting out and appear impervious to punishment. Additional difficulties might include externalizing behavior problems (lying, cheating, stealing, temper outbursts, jealousy) and school

dropout. Boys frequently report physical abuse and having run away; girls similarly report physical abuse but also having been sexually abused. These boys and girls are also likely to be sexually active. Often they are not particularly motivated to become involved in therapy. Because Scale 4 is quite heterogeneous with a correspondingly high number of descriptors, a formal scoring and inspection of the Harris-Lingoes scales can often be extremely useful in determining which of the scale descriptors is most appropriate.

Scale 5. Masculinity-Femininity (*Mf*)

This scale was originally designed to identify males who had homosexual tendencies and gender-identity confusion. However, it has been largely unsuccessful because a high score does not seem to clearly and necessarily relate to a person's sexual orientation. Instead, it relates to the degree to which a person endorses items related to traditional masculine or feminine roles or interests. Males who have completed university degrees usually score 5 *T* scores higher ($T = 60-65$) than the standardization sample; and those with less than a high school education will score, on average, 5 *T* scores lower. Interpretations, therefore, should consider the influence of education. In contrast, the correlation between females and education is quite small ($-.15$ correlation). The item content seems to be organized around these five dimensions: personal and emotional stability, sexual identification, altruism, feminine occupational identification, and denial of masculine occupations. The items are scored in the opposite direction for females. Thus, high scores for males have traditionally been used to suggest a nonidentification with stereotyped masculine interests, whereas a high score for females has traditionally been used to suggest an identification with these masculine interests.

An important consideration regarding Scale 5 is that it is not an actual clinical scale in the same sense as most of the other scales. It does not actually assess any pathological syndromes and thus does not provide clinical information. As a result, a "treatment implications" section has not been included. However, it can be useful in providing color or tone to the other scales. Interpretations should first be made of the other scales, and then the meaning of the relative score on Scale 5 should be taken into consideration. For example, an elevation on Scale 4 (*Pd*) would indicate that the person is impulsive, might act out under stress, and feels alienated from him- or herself or society. If the person scoring a high 4 is a male and also scores low on Scale 5, he would be likely to express his dissatisfaction through action, have low insight into his behavior, and place emphasis on physical strength. In contrast, a male with a high scale on 4 accompanied by a high score on 5 suggests that he will be more introspective, sensitive, and articulate and may channel his antisocial feelings toward creating social change. As noted previously, the person's level of education and socioeconomic status should be taken into account when deciding whether the score is "high" or "low."

High Scores on Scale 5 (Males)

- Aesthetic and artistic interests.
- Little interest in stereotypically masculine interests.

- More likely than most men to be involved in child rearing and housekeeping.
- Insightful, sensitive, and introspective (qualities conducive to benefiting from psychotherapy).
- Reduced likelihood that any existing pathology will be acted out.

Low Scores on Scale 5 (Males)

- Stereotypically male interests, occupations, hobbies, and other activities.
- Presenting themselves as extremely “masculine.”

High Scores on Scale 5 (Females)

- Interest in traditionally masculine interests and activities.
- Involved in occupations that are more frequently occupied by males.
- Possible difficulty engaging in traditional psychotherapy because they usually do not value introspection and insight, might have difficulty articulating their problems and expressing emotions.

Low Scores on Scale 5 (Females)

- Endorsement of many traditionally feminine roles, behaviors, and interests.
- Considerable satisfaction is likely to be derived from involvement as mothers and spouses.

MMPI-A Considerations

Scale 5 elevations for males were rare on both the MMPI-A clinical and normative samples. Males who do score high will seem interested in stereotypically feminine interests, deny stereotypically masculine interests, and are less likely to act out. If there are correspondingly high elevations on other scales suggesting acting out (Scales 4, 9, *F*), these should be given more consideration than the suppression value of an elevated Scale 5. Further research needs to be conducted on the behavioral correlates of both high- and low-scoring adolescent females. However, tentative interpretations would indicate that high-scoring females have stereotypically masculine interests.

Scale 6. Paranoia (*Pa*)

Scale 6 was designed to identify persons with paranoid conditions or paranoid states. It measures a person's degree of interpersonal sensitivity, self-righteousness, and suspiciousness. Many of the 40 items center on areas such as ideas of reference, delusional beliefs, pervasive suspiciousness, feelings of persecution, grandiose self-beliefs, and interpersonal rigidity. Whereas some of the items deal with overt psychotic content, other less extreme questions ask information related to the perceived ulterior motives of others. The Harris-Lingoes subscales divide the items in Scale 6 into ideas of external influence, poignancy (feelings of being high strung, sensitive, having stronger feelings than others, and a sense of interpersonal distance), and naiveté (overly optimistic, high morality, denial of hostility, overly trusting, and vulnerability to being hurt).

Mild elevations on Scale 6 suggest that the person is emotional, softhearted, and experiences interpersonal sensitivity. As the elevation increases, a person's

suspicion and sensitivity become progressively more extreme and consistent with psychotic processes. He or she may have delusions, ideas of self-reference, a grandiose self-concept, and disordered thought processes. In contrast, low-scoring persons are seen as being quite balanced. However, there are some differences between the descriptions given for low-scoring males and low-scoring females. Low-scoring males are described as cheerful, decisive, self-centered, lacking in a strong sense of conscience, and having a narrow range of interests. Females are somewhat differently described as mature and reasonable.

In some ways, Scale 6 is quite accurate in that high-scoring persons usually have significant levels of paranoia. However, the contents of most of the 40 items are fairly obvious. Thus, a person wanting to conceal his or her paranoia, because of fear over the imagined consequences of detection, could do so quite easily. Because of this fact, it might be possible for low or moderate scores to still be consistent with paranoia. This is especially true for bright and psychologically sophisticated persons. They might mask their paranoia not only on the test but also in real life. They might be a member of some extreme political group or religious cult that provides some degree of social support for their underlying paranoid processes. However, if the scale is clearly elevated, it is an excellent indication of paranoia.

High Scores on Scale 6

- Highly suspicious, vengeful, brooding, resentful, and angry.
- Will feel mistreated and typically misinterpret the motives of others, feeling that they have not received a fair deal in life.
- May have a thought disorder with accompanying ideas of reference, delusional thinking, fixed obsessions, compulsions, and phobias.
- Extremely rigid thinking, argumentative.
- May easily misinterpret the benign statements of others as personal criticisms.
- Will enlarge on and brood over partially or wholly invented criticisms.
- Underlying feelings of anger expressed in a rigidly moralistic and intellectual manner.
- Will reduce anxiety through intellectualization.
- Use projection to deny underlying feelings of hostility.
- Personal hostility might be expressed through indirect means (yet appear outwardly self-punishing).
- Will feel as if they have gotten an unfair deal from life, will resent family members.
- Moderate elevations: much less likely to reflect overtly psychotic trends but will still be suspicious, argumentative, potentially hostile, and quite sensitive in interpersonal relationships.
- Mild elevations (for nonpatient groups) are usually described in relatively favorable terms: hardworking, industrious, moralistic, sentimental, softhearted, peaceable, generous, trusting unless betrayed, intelligent, poised, rational, fair-minded, and with a broad range of interests.
- Mild elevations (for nonpatient groups) may also be submissive, prone to worry, high strung, dependent, and lacking in self-confidence.

- Mild elevations (for psychiatric patients): oversensitive, slightly paranoid, suspicious, feel as if their environment is not sufficiently supportive.

Related Scale Elevations

- *High 6 and 8* (see 68/86 code type). Highly suggestive of paranoid schizophrenia.
- *Corresponding elevation on Scale 3* (see 36/63 code type). Will repress their hostile and aggressive feelings and appear naive, positive, and accepting; may easily enter into superficial relationships, but after these relationships deepen, their underlying suspiciousness, hostility, ruthlessness, and egocentricity will become more openly expressed.

Treatment Implications

- Provides an index of the degree to which clients can develop a trusting relationship, their attitudes toward authority figures, and their degree of flexibility.
- Psychotherapy may be extremely difficult because of their rigidity, poor level of insight, and suspiciousness (check TRT/Negative Treatment Indicators Scale).
- Are often argumentative, cynical, and resentful, thereby making it difficult to establish a relationship of mutual trust, empathy, and respect.
- Do not like to discuss emotional issues, overvalue rationality, and are likely to blame others for their difficulties.
- Frequently will not return following the initial session; will leave feeling that they have not been understood. Major challenge with an intake is to make sure that they feel understood (see the next “Client Feedback Statements”).
- Might attempt to manipulate the therapist by implicitly suggesting they will terminate.
- Very high scores (check *BIZ*/Bizarre Mentation and critical item clusters related to mental confusion and persecutory ideas subscales) may require medication.
- Evaluate potential for dangerousness toward others if brooding and resentment are particularly pronounced.

Client Feedback Statements[†]

You are rational, fair-minded, and loyal. You have high personal standards, and you work hard to be above criticism or judgment. You have strong values, and you may be very black-and-white about the right and wrong way of seeing and doing things You are susceptible to anything that can be construed as criticism or judgment. Currently you may feel on edge and tense, as if someone is going to unfairly criticize or attack you. At times, your sensitivity can shade toward paranoia so that it is hard for you to know whom to trust. These times might be quite frightening because you don't know whether your mistrust of others is due to your sensitivity or whether you are truly seeing things clearly If you feel unfairly treated or if you feel others are mistreated, it makes you angry, and you feel driven to “right the wrong.” If people hurt you, even though you may forgive them, it is hard for you to forget what they have done You tend not to let others know when you are hurt or angry until you feel you are completely justified in doing so. However, by that time you are angry and have hard feelings about the other person not being sensitive to your needs. You may store and rationalize your resentments without

letting people know how you feel, and if they continue their actions you begrudge them and feel better. You don't ask for what you want until you feel you fully deserve it. Because you are so sensitive to the issue of fairness, when you finally express your feelings you may try to explain why you are feeling hurt or angry. This is your way of justifying yourself; however, it makes others feel defensive, so they tend to argue back.

MMPI-A Considerations

Elevations are consistent with academic problems, including poor grades and suspension (check A-sch/School Problems content scale). Clinical girls report significant disagreements with their parents (check A-fam/Family Problems content scale). Clinical boys are described as hostile, withdrawn, immature, and argumentative; they feel persecuted and are not well liked by their peers. In addition, they are perceived as being overly dependent on adults, attention-seeking, resentful, anxious, and obsessed; they feel as if they are bad and deserving of punishment. Because the items on the MMPI-2 and MMPI-A are the same, the Harris-Lingoes scales can be used to understand possible patterns of item endorsement.

Scale 7. Psychasthenia (*Pt*)

The 48 items on Scale 7 were originally designed to measure the syndrome of psychasthenia. Although psychasthenia is no longer used as a diagnosis, it was current when the MMPI was first developed. It consisted of compulsions, obsessions, unreasonable fears, and excessive doubts. Thus, it is quite similar to what today would be an anxiety disorder with obsessive-compulsive features. However, there are important differences between Scale 7 and obsessive-compulsive disorder. Scale 7 measures more overt fears and anxieties that the person might be experiencing (check also ANX/Anxiety). In contrast, persons struggling with obsessive-compulsive disorder could potentially score quite low on 7 because their behaviors and obsessions are effective in reducing their levels of anxiety (check the OBS/Obsessiveness content scale). Although an elevation on Scale 7 may suggest the possibility of obsessive-compulsive disorder, other anxiety-related disorders or situational states could also produce an elevation.

Scale 7 is the clinical scale that most clearly measures anxiety and ruminative self-doubt. Thus, along with elevations on Scale 2, it is a good general indicator of the degree of distress the person is currently undergoing. High scorers are likely to be tense, indecisive, and obsessively worried and to have difficulty concentrating. In a medical context, they are prone to overreact to even minor medical complaints. They are usually rigid, agitated, fearful, and anxious. The most frequent complaints will be related to cardiac problems as well as difficulties related to their gastrointestinal or genitourinary systems. In nonmedical and more normal populations, high scorers are likely to be high strung, articulate, individualistic, and perfectionistic, with extremely high standards of morality.

High Scores on Scale 7

- Apprehensive, worrying, perfectionistic, tense, difficulty concentrating.
- Highly introspective, self-critical, self-conscious, and feel a generalized sense of guilt.

- Wide variety of superstitious fears.
- Orderly, conscientious, reliable, persistent, organized.
- Often moralistic with high standards for themselves and others.
- Lack in originality.
- Even minor problems might become sources of considerable concern.
- Will overreact to and exaggerate the importance of events.
- Will use rationalization and intellectualization to reduce anxiety; these defenses are rarely successful.
- Defenses against their anxiety could be expressed in a variety of rituals.
- Experience self-doubt, rigid, meticulous, apprehensive, uncertain, and indecisive.
- Social difficulties, frequently worrying about their degree of acceptance and popularity.
- Extremely high scores: disruption in a person's ability to perform daily activities.

Related Scale Elevations

- *Moderate elevation on 7 and 2* (see 27/72 code type). Good prognosis for therapy because clients are sufficiently uncomfortable to be motivated to change.
- *Elevations on Scales 7 and 8, with Scale 7 relatively higher* (10 *T* score points or more) *than 8* (see 78/87 code type). Person is anxious about and struggling with an underlying psychotic process; better prognosis.
- *Elevations on Scales 7 and 8, with Scale 7 relatively lower than 8* (10 *T* score points or more). Person is likely to have given up attempting to fight the disorder, psychotic processes are either of a chronic nature or likely to become more chronic; poor prognosis.

Treatment Implications

- Usually highly motivated to change, will usually stay in therapy, progress tends to be slow but steady.
- Immediate task is to work directly with their anxiety (e.g., using cognitive restructuring, hypnosis, relaxation, or systematic desensitization).
- Anxiety may be sufficiently high that antianxiety medication may be indicated. This should be considered to help clients work more constructively in a therapeutic context and function in their daily activities.
- Insight-oriented therapy should be used with caution; these clients have a tendency to intellectualize and ruminate indefinitely without making any concrete changes and may be overly perfectionistic and rigid, thereby making it difficult for them either to accept insights or to integrate them in a flexible, problem-solving manner.

Client Feedback Statements[†]

You are a thoughtful, analytical, responsible individual who takes life seriously. You tend to be detail oriented, reliable, and thorough. You generally follow the rules and are trustworthy. Dutiful and conscientious, you're the kind of person people can count on Currently, however, some of your strengths may be working against you. You spend a lot

of time standing back and observing the world, worried that some mistake you make will lead to catastrophe. You often feel a sense of anxiety, as if something bad is about to happen. Even when things are going well, it's hard for you to switch off your mind to get rid of that feeling of nervousness It's hard for you to be spontaneous because you see every side of every issue. When you have to make a decision, you are likely to overanalyze, worrying that you might have missed some important detail. Part of the reason you fret so much about making a mistake is that when something goes wrong you feel so deeply guilty. Not only do you focus on possible future mistakes, but you also spend a lot of time thinking about the past, about oversights you've made, and obsessing about how much guilt you should feel You tend to be your own worst critic. Even when things go well, you can't relax and enjoy them.

MMPI-A Considerations

Few descriptors have been found for adolescents with high scores on *Pt*, in part because, it is speculated, an early (adolescent) rigid personality style may not become problematic until later in adult life. Girls from clinical populations are likely to be depressed, may make suicidal threats, are more likely to steal, and report significant disagreements with their parents. Boys from clinical populations are likely to have low self-confidence and may have been sexually abused. However, more research needs to be performed to more clearly understand the behavioral correlates of adolescents who score high on Scale 7.

Scale 8. Schizophrenia (*Sc*)

Scale 8 was originally designed to identify persons who were experiencing schizophrenic or schizophrenic-like conditions. This goal has been partially successful in that a diagnosis of schizophrenia is raised as a possibility in the case of persons who score extremely high. However, even persons scoring quite high would not necessarily fulfill the criteria for schizophrenia, in part because the items in the scale cover a highly diverse number of areas. Thus, elevations can occur for a variety of reasons, which means that the descriptions of high scorers are also quite varied. The items assess areas such as social alienation, apathy, poor family relations, unusual thought processes, and peculiarities in perception. Other questions measure reduced efficiency, difficulties in concentration, general fears and worries, difficulty coping, and problems delaying impulses. Because of the many scale items, heterogeneity of their content, and the resulting numerous potential descriptors for individuals scoring high on Scale 8, it can be useful to consult the Harris-Lingoes subscales to more fully understand the meanings of elevations. Harris and Lingoes (1968) described the next six different content areas.

1. Social alienation
2. Emotional alienation
3. Lack of ego mastery—cognitive (strange thought processes, fear of losing his or her mind, difficulty concentrating, feelings of unreality)
4. Lack of ego mastery—conative (difficulty coping with everyday life, low interest in life, hopelessness, depression)

5. Lack of ego mastery—defective inhibition (impulsive, hyperactive, sense of being out of control, impulsive, laughing or crying spells)
6. Bizarre sensory experiences

In general, an elevated score on Scale 8 suggests the person feels alienated, distant from social situations, and misunderstood. He or she might have a highly varied fantasy life and, when under stress, will withdraw further into fantasy. Others will most likely perceive the person as eccentric, seclusive, secretive, and inaccessible. He or she often has a difficult time maintaining a clear and coherent line of thought. Communication skills will be poor; other people often feel they are missing some important component of what this individual is trying to say. The person will typically not make clear and direct statements and often has difficulty focusing on one idea for very long.

Adolescents score higher on Scale 8, which might be consistent with their greater openness to unusual experiences, turmoil in establishing a solid sense of identity, and greater feelings of alienation. Some groups of relatively normal persons might have mild elevations on 8. These might include individuals developing sensory impairments, persons with organic brain disorders, or unconventional persons who identify with the counterculture. Persons who have had a variety of drug experiences may score somewhat higher on 8. This may reflect the direct effects of the drugs themselves rather than suggest greater levels of pathology.

High Scores on Scale 8

- Unusual beliefs, unconventional.
- Possible difficulty concentrating and focusing attention.
- *Moderately elevated.* Merely aloof, different, approach tasks from an innovative perspective; may have philosophical, religious, or abstract interests; little concern with concrete matters.
- *Moderately elevated.* Described by others as shy, aloof, and reserved.
- *Higher elevations.* Greater difficulties organizing and directing thoughts, aggressive, resentful, and/or hostile feelings yet cannot express these feelings.
- Positive qualities might include being peaceable, generous, sentimental, sharp-witted, interesting, creative, and imaginative.
- *Very high elevations.* Bizarre mentation, delusions, highly eccentric behaviors, poor contact with reality, and possibly hallucinations (see BIZ/Bizarre Mentation content scale), feel incompetent, inadequate, and plagued by a wide variety of sexual preoccupations, self-doubts, and unusual beliefs.
- *Very high scores.* Reflect unusual experiences reported by extremely anxious patients, adolescent adjustment reactions, pre-psychotic disorders, borderline personalities, or relatively well-adjusted persons who are malingering.

Related Scale Elevations

- *Elevations on 4 and 8* (see 48/84 code type). Extremely distrustful, alienated from their world, environment perceived as dangerous, likely to react to others in a hostile and aggressive fashion.

- *High 8 and 9* (see 89/98 code type). Likely to constantly deflect the direction of conversation, frequently diverting it to unusual tangents, distorted view of their world; and individuals have the energy to act on these distorted perceptions.
- *Elevated 8 and F, 2, 4, and 0*. Schizoid profile.

Treatment Implications

- *Very high scores*. Difficulty trusting others and developing relationships resulting in difficulty with psychotherapy, especially during its initial stages.
- After initial therapy, individuals often stay in therapy longer than many other types of clients; may eventually develop a relatively close and trusting client/therapist relationship.
- Treatment should be focused around working on specific current problems the client is dealing with.
- Often prognosis is poor due to the often-chronic nature of clients' difficulties.
- With extremely disorganized thought processes, referral for medication might be indicated.

Client Feedback Statements[†]

You are an imaginative, creative person who thinks differently from others Currently, you seem to be knocked off balance and confused. The world may be a somewhat frightening place right now because it's hard for you to read people and to know how they're feeling toward you. You may feel disconnected from others, as if you are looking at the world from a distance Even when things are going well or in a tender or sweet moment, you may find yourself feeling strangely cold or even angry or disgusted. Moments that others find happy or tender might leave you untouched. You may experience dark moods where you suddenly feel angry, empty, and irritable, and you might not know where the mood comes from. In fact, these dark moods may sweep over you even when things are going relatively well You have a tendency to daydream and spend time inside your mind fantasizing; sometimes the daydreams may be disturbing. Spending time in your own thoughts may make it hard for you to get things done. Life must feel somewhat gray, empty, and at times meaningless. It's hard to get motivated and to have goals and ambitions because nothing seems worthwhile or rewarding.

MMPI-A Considerations

Both boys and girls elevated on this scale report a higher rate of having multiple school-related problems, with boys frequently being suspended and girls being unlikely to report having had any significant achievements (check A-sch/School Problems content scale). In addition, the possibility of sexual abuse should be investigated. Girls are likely to report increased disagreements with their parents (check A-fam/Family Problems content scale) and, among clinical populations, may be aggressive, threaten suicide, act out, and have outbursts of temper. In contrast, clinical boys are described as having behaviors such as being guilt-prone, shy, withdrawn, fearful, and perfectionistic; showing low self-esteem; being "clingy"; and having somatic complaints (e.g., nausea, headaches, dizziness, stomach pains). Clinical boys with quite high elevations

might also have psychotic features, including delusions, hallucinations, ideas of reference, grandiose beliefs, or peculiar speech and mannerisms (check A-biz/Bizarre Mentation content scale).

Scale 9. Hypomania (*Ma*)

The 46 items on Scale 9 were originally developed to identify persons experiencing hypomanic symptoms. These symptoms might include cyclical periods of euphoria, increased irritability, and excessive unproductive activity that might be used as a distraction to stave off an impending depression. Thus, the items are centered on topics such as energy level, irritability, egotism, and expansiveness. The Harris-Lingoes subscales classify the content of the items under amorality, psychomotor acceleration, imperturbability, and ego inflation. However, hypomania occurs in cycles. Thus, persons in the acute phase were unable to be tested because of the seriousness of their condition. Further, some persons might score quite low on Scale 9, which might reflect the depressive side of their cycle. These low scorers, then, might still develop a hypomanic state and may have actually been hypomanic in the past.

The scale is effective not only in identifying persons with moderate manic conditions (extreme manic patients are generally untestable) but also in identifying characteristics of nonpatient groups. A full 10% to 15% of normals have elevations on this scale, suggesting characteristics such as an unusually high drive level. Males with mild to moderate elevations and with no history of psychiatric disturbance might be described as warm, enthusiastic, outgoing, and uninhibited. They would most likely be able to expend a considerable amount of energy over a sustained period of time. They might also be easily offended, hyperactive, tense, and prone to periods of worry, anxiety, and depression. Others might describe them as expressive, individualistic, generous, and affectionate. Nonpatient females are likely to be frank, courageous, talkative, enthusiastic, idealistic, and versatile. Their friends and partners are likely to describe them as making big plans, wearing strange or unusual clothes, stirring up excitement, becoming very excited for no reason, being risk takers, and telling people off. High-scoring males were described by others as demanding excessive attention, being bossy, talking back to others without thinking, whining, and taking nonprescription drugs.

Age and race are important when evaluating what should be considered a high or low score. Some studies have indicated that certain populations of African Americans score higher than European Americans. Also, younger populations (adolescents and college-age students) score somewhat higher than nonpatient adults. In contrast, older adults often score quite low on Scale 9.

High Scores on Scale 9

- *Extremely high scores.* Moderate manic episode, will be maladaptively hyperactive, poorly focused, flighty ideas, inflated sense of self-importance, and low impulse control.
- Possibly perceived as creative, enterprising, and ingenious, but their view of what they can actually accomplish is unrealistic.
- Unwarranted sense of optimism.

- Become irritable with relatively minor interruptions and delays.
- Expend a considerable amount of energy, but their activity is usually unproductive because it is unfocused.
- Increased energy may serve to distract them from painful feelings or situations.
- Possible good initial impression because they are enthusiastic, friendly, and pleasant, but also deceptive, manipulative, and unreliable, ultimately causing interpersonal difficulties.
- Quickly develop relationships with others, but these relationships will be superficial.
- Will be perceived as restless and agitated.
- *Moderate elevations.* Often more able to focus and direct their energy in productive directions.
- *Moderate elevations among nonpatients.* Direct, energetic, enthusiastic, sociable, independent, optimistic, have a wide range of interests; might also be somewhat guileful, overactive, impulsive, persuasive, and prefer action to thought.
- Might sometimes show mood difficulties and experience elation without cause.
- Self-centered and impulsive.
- Scores alone not sufficient to distinguish a person who is energetic, optimistic, and focused from a person who is scattered, ineffective, and hyperactive (note critical items, Harris-Lingoes subscales, integrate relevant historical information).

Related Scale Elevations

- *High 9 and 2* (NB: These are usually negatively correlated). Reflects an agitated state, person is attempting to defend self from underlying hostile and aggressive impulses, might be highly introspective and narcissistically self-absorbed. 9 and 2 can also be elevated for patients with certain types of organic impairment.
- *High 9 with low 2 and 7.* Suggests a minimum of psychological distress. Males are likely to have a compulsive need to seek power and place themselves in narcissistically competitive situations. With elevated *K*, these males are likely to be managerial, autocratic, and power hungry and to expend a considerable degree of effort organizing others. Self-esteem would often be dependent on eliciting submission and weakness from others (what they usually receive is a grudging deference rather than admiration). Females having this profile are likely to be prone to exhibitionistic self-display and to be extremely concerned with their physical attractiveness.
- *Low scores ($T < 41$).* Likely to be apathetic, depressed, fatigued, pessimistic, feel inadequate.

Treatment Implications

- Distractibility and overactivity often make psychotherapy difficult. Clients resist focusing on problems by diverging onto irrelevant tangents and object to psychological interpretations of their behavior.
- May develop grandiose plans for change, but these are seldom followed through.

- Tend to use denial and avoid self-examination.
- Low frustration for tolerance and frequent irritability might result in dramatic therapy sessions.
- Frequent disregard for prearranged appointment times, cancel because they are too busy.
- High resistance indicates they might optimally benefit from non- or self-directive interventions or paradoxical strategies.
- Evaluate for the possibility of a bipolar disorder with follow-up for appropriate medication.
- Check for alcohol or drug abuse (check MAC-R, AAS/Addiction Acknowledgment scale, and APS/Addiction Potential scale).
- *Low score on 9.* Difficulty becoming motivated, require a concrete action program with a high degree of structure.

Client Feedback Statements[†]

You are an energetic, driven, ambitious individual who is able to think and move quickly and get a lot done You may have periods where you're so optimistic that you over-commit and take on so many tasks and activities that it is impossible to complete them all. Sometimes your energy may be so high that you have a reduced need for sleep and you feel impatient and angry with a world that moves too slowly. During these periods your moods may quickly shift from positive to negative; you may be upbeat and cheerful one moment and in the next angry and irritated, feeling that people are blocking you from getting what you want You may feel so much energy that you can become irritated with people for not keeping up with you. When your energy is high, working on a single task can be difficult. You have a tendency to see the connections between things, so it's easy for you to become distracted and sidetracked. At other times, you may become so focused on a particular idea or activity that you keep at it when others feel you should let it go. At these times, when people try to persuade you that you are being unrealistic you can become quite angry, even explosive and aggressive.

MMPI-A Considerations

Moderate elevations suggest that the individual is enthusiastic, animated, and takes an interest in things. However, higher elevations suggest underachievement in school and problems at home (check A-sch/School Problems and A-fam/Family Problems content scales). Scale 9 elevations might also reflect irrational, manic behaviors and antisocial acts (check A-con/Conduct Problems content scale). Among boys, amphetamine use is relatively common. Adolescents with elevations on this scale are typically insensitive to criticism, do not like to reflect on their behavior, and are therefore unmotivated to become involved in therapy. They may also believe that they know more than authority figures and feel that such persons punish people unjustly. They might be self-confident, oppositional, take advantage of others, and deny any social discomfort.

Scale 0. Social Introversion (Si)

This scale was developed from the responses of college students on questions relating to an introversion-extraversion continuum. It was validated based on the degree

to which the students participated in social activities. High scores suggest that the respondent is shy, has limited social skills, feels uncomfortable in social interactions, and withdraws from many interpersonal situations. Such persons would prefer to be alone or with a few close friends than with a large group. One cluster of items deals with self-depreciation and neurotic maladjustment; the other group deals with the degree to which the person participates in interpersonal interactions. The different item contents have been further organized around the areas of shyness/self-consciousness, social avoidance, and the extent that a person feels alienated from self and others (Ben-Porath, Hostetler, Butcher, & Graham, 1989). These contents form subscales that can be used in conjunction with the Harris–Lingoes subscales to help determine the different variables related to why a person had an elevation on Scale 0 (see section on “MMPI-2 Harris–Lingoes and *Si* Subscales” later in this chapter).

Scale 0 is similar to 5 in that it is used to “color” or provide a different emphasis to the other clinical scales. Thus, interpretations should first be made without considering 5 and 0; later, the implications of these scales should be included. As a result, code types involving 0 have not been included in the section on 2-point codes. Elevations on 0 help provide information on the other scales by indicating how comfortable persons are with interactions, their degree of overt involvement with others, the effectiveness of their social skills (check SOD/Social Discomfort content scale), and the likelihood that they will have a well-developed social support system. A low score on 0 often reduces the degree of pathology that might otherwise be suggested by elevations on the other scales. A low 0 also suggests that, even if persons have a certain level of pathology, they are able to find socially acceptable outlets for these difficulties. In contrast, a high 0 suggests an exaggeration of difficulties indicated by the other scales. This is particularly true if 0, 2, and 8 are all elevated. This finding suggests that the person feels socially alienated, is withdrawn, is self-critical, and has unusual thoughts. However, he or she is not likely to have an adequate social support group to help in overcoming these difficulties. Although an elevated 0 can suggest an increase in personal difficulties, it often reflects a decreased likelihood of acting out. This finding is further supported by corresponding elevations on 2 and 5 (for males or a lowering for females). As a result, 0, 2, and 5 are often referred to as *inhibitory scales*.

High Scores on Scale 0

- Feel uncomfortable in group interactions, may have poorly developed social skills.
- Self-effacing, lacking in self-confidence, submissive, shy, timid.
- Others might experience them as cold, distant, rigid, and difficult to get to know.
- *Extremely high scorers*. Withdrawn, ruminative, indecisive, insecure, retiring, uncomfortable regarding their lack of interaction with others, sensitive to the judgments others make of them.
- Will not have a well-developed social support group to help them overcome difficulties.
- *Moderate scores*. Dependable, conservative, cautious, unoriginal, serious, over-controlled.
- *Normal males who score high on 0*. Modest, inhibited, lacking in self-confidence, generally deficient in social presence.
- *Normal females who score moderately high*. Similarly described as modest, shy, self-effacing, sensitive, and prone to worry.

Low Scores on Scale 0

- Warm, outgoing, assertive, self-confident, verbally fluent, gregarious.
- Strong need to be around other people.
- Concerned with power, recognition, status.
- May be opportunistic, exhibitionistic, manipulative, and self-indulgent; in extreme cases they might also be immature, self-indulgent, superficial.
- *Normal males.* Sociable, expressive, socially competitive, verbally fluent.
- *Normal females.* Sociable, talkative, assertive, enthusiastic, adventurous.
- *Extremely low scores.* May have highly developed social techniques but behind their external image, they may have feelings of insecurity with a strong need for social approval, hypersensitive, difficulties dealing with feelings of dependency, large number of superficial friends but probably do not feel close to anyone.

Treatment Implications

- *High scorers* (extremely introverted). Difficulty engaging in therapy because they are shy, withdrawn, and anxious; would take time to develop a therapeutic relationship; may expect the therapist to be directive and dominate. A withdrawn and nondirective therapist might increase such a client's anxiety, resulting in premature termination.
- Might appear unmotivated and passive but internally they are likely to feel high strung and anxious (check LSE/Low Self-Esteem, A-lse/Low Self-Esteem, SOD/Social Discomfort, and A-sod/Social Discomfort content scales).
- Likely to be overcontrolled and experience considerable difficulties in making changes.
- Group treatment and social skills training are often appropriate interventions. It is essential that the group should be supportive and accepting, thereby increasing the likelihood that clients would experiment with new behaviors.
- *Low scorers* (extremely extraverted). Possible difficulty due to superficial orientation, disinclination to reflect inwardly.

MMPI-A Considerations

Among adolescents, high scores on 0 are a clear indication of difficulties in social relationships, particularly related to low self-esteem and social withdrawal. The behavioral correlates for girls suggest that they are withdrawn, shy, fearful, depressed, may have had suicidal ideation and/or gestures, have eating problems, are socially withdrawn, and have only a few friends. Elevations also suggest an inhibitory effect in that they are unlikely actually to act out on their pathology. Thus, they rarely report difficulties with drugs or alcohol, delinquency, or sexual acting out and have little interest in dating or sexual relationships. There are less behavioral correlates for boys, but high scores do suggest that they are unlikely to participate in school activities.

MMPI-2 2-POINT CODES

Code-type interpretation often produces more accurate and clinically useful interpretations than merely interpreting individual scales. The basis of code-type interpretation

depends on empirical correlations among various classes of nontest behavior. The 2-point codes included in the following section have been selected based on their frequency of occurrence, the thoroughness of the research performed on them, and their relative clinical importance. Thus, some combinations of code types will not be discussed.

Code-type interpretation is most appropriate for disturbed populations in which *T* score elevations are at least 65 on the MMPI-2 or MMPI-A. The descriptions are clearly oriented around the pathological dimensions of an individual. The 2-point code descriptions, then, do not have the same divisions into low, moderate, and high elevations as the individual scores but are directed primarily toward discussions of high elevations. When considering 2-point codes that are in the moderate range ($T = 60-70$), interpretations should be made with caution, and the more extreme descriptions should be considerably modified or excluded.

Usually, the elevation of one scale in relation to the other does not make much difference as long as the elevations are still somewhat similar in magnitude. A general approach is that if one scale is 10 points or more higher than the other, the higher one gives more color to, or provides more emphasis for, the interpretation. Specific elaborations are made for scales in which a significant difference between their relative elevations is especially important. If the scales have an equal (or nearly equal) magnitude, they should be given equal emphasis.

In some cases, more than two scales will be equally elevated, thereby making it difficult to clearly establish which scales represent the 2-point code. In these cases, clinicians should look at the descriptions provided for other possible combinations. For example, if Scales 2, 7, and 8 are elevated for a particular profile, the clinician should look up the 27/72 code as well as codes 78/87 and 28/82. The descriptions for the code type with the highest elevations and those descriptors that are common between the different code descriptions are most likely to be valid. However, multiple elevations also raise the issue of the generalizability of the MMPI descriptors (which the majority of research has been derived from) and the MMPI-2 (Butcher & Williams, 2000; D. Edwards et al., 1993; Humphrey & Dahlstrom, 1995; Tellegen & Ben-Porath, 1993). Up to 50% of the code types have been found to differ, which is particularly true for poorly defined code types. This fact potentially compromises the validity of the code-type descriptions. A more cautious approach would be to rely more on the single-scale descriptors.

In developing meaningful interpretations, it is important to continually consider the underlying significance of elevated scales. Doing this requires considering factors such as the manner in which the scales interact, the particular category of psychopathology they suggest, and their recurring patterns or themes. When possible, *DSM-5* classifications have been used, but the term *neurosis* is used occasionally because of its ability to summarize a wide variety of disorders and/or its ability to refer to a cluster of related scales (e.g., “neurotic triad”). Some characteristics described in the code types will be highly accurate for a specific person whereas others will not be particularly relevant or accurate. Clinicians, then, will need to continually reflect on their data to develop descriptions and diagnoses that are both accurate and relevant.

The code types from the MMPI-A should be used with considerable caution because there is currently insufficient research on the behavioral correlates of these code types. In contrast, there is considerable research on the correlates of individual MMPI-A

scale elevations. With this caution in mind, it is recommended that clinicians tentatively use the code types described in the following pages to help generate hypotheses concerning adolescent functioning. Doing this is partially justified in that many of the MMPI code type correlates are common for both adults and adolescents (Archer, 1992a). In addition, the majority of the code types derived from the MMPI will be the same for the MMPI-A, especially if the code types are well defined.

12/21

Symptoms and Behaviors

- Complaints revolve around physical symptoms; can be either organic or functional (check the HEA/Health Concerns content scale).
- Common complaints: pain, irritability, anxiety, physical tension, fatigue, over-concern with physical functions.
- Significant level of depression is present.
- Characteristically handle psychological conflict through repression and by attending to real, exaggerated, or imagined physical difficulties.
- Even if physical difficulties are organically based, symptoms will be exaggerated and used to manipulate others. These individuals elaborate their complaints beyond what can be physically confirmed, often doing so by misinterpreting normal bodily functions.
- Have learned to live with their complaints and use them to achieve their own needs.
- More frequently encountered in males and older persons.
- *Pattern 1* (generalized hypochondriac). Significant depressive features, self-critical, indirect, manipulative; if difficulties are solely functional, they are more likely to be shy and withdrawn. Persons with a significant organic component are likely to be loud complainers. Complaints are usually focused around the trunk of the body and involve the viscera (in contrast to the 13/31 code in which complaints are more likely to involve the central nervous system and peripheral limbs; see 13/31 code type).
- *Pattern 2* (chronic pain patient). Have given in to their pain and learned to live with it. Expression of pain is likely to be exaggerated, used to manipulate others. Check for history of drug or alcohol abuse as means of self-medication.
- *Pattern 3* (patients with recent, severe accident). Their elevations on Scales 1 and 2 reflect an acute, reactive depression that occurs in response to the limiting effects of their condition.
- Some heavy drinkers have elevations on 1, 2, 3, and 4; they experience considerable physical discomfort, digestive difficulties, tension, depression, hostility, poor work and relationship histories.

Personality and Interpersonal Characteristics

- Introverted, shy, self-conscious, passive dependent.
- May harbor resentment against persons for not providing them with sufficient attention and emotional support.
- Interpersonally sensitive and manipulate others with their symptoms.

Treatment Implications

- Lack insight, not psychologically sophisticated, resent any implications that their difficulties may be even partially psychological (check the TRT/Negative Treatment Indicators scale).
- Difficult to take responsibility for their behavior.
- Typically seek medical explanations and solutions for their difficulties.
- Somatize stress, thus they are able to tolerate high levels of discomfort before being motivated to change.
- Generally not good candidates for psychotherapy, especially if the therapy is insight oriented.

13/31

Symptoms and Behaviors

- *Classic conversion V* (Scale 2 is 10 points or more lower than 1 or 3). As 2 becomes lower in relation to 1 and 3, the likelihood of a Conversion Disorder increases, strengthened in males who have correspondingly high Scales 4 and 5 and in females with a correspondingly high 4 but lowered 5).
- 13/31 is more frequent in females and older adults than in males and younger persons.
- Very little anxiety due to conversion into physical complaints (check corresponding elevations of Scales 2 and 7 as anxiety is sometimes present).
- *If anxiety/depression are present*. Conversions are currently unable to eliminate their conflicts effectively.
- Extensive complaining about physical difficulties.
- Complaints typically involve problems related to eating, such as obesity, nausea, anorexia nervosa, or bulimia; also psychogenic seizures or psychogenic paralysis.
- Possible presence of vague “neurological” difficulties, such as dizziness, numbness, weakness, and fatigue.
- Occasional sense of indifference (marked lack of concern) regarding symptoms.
- Strong need to appear rational and socially acceptable, yet nonetheless control others through histrionic and symptom-related means.
- Often try and appear psychologically hypernormal (check *K* and *L*).
- Even if complaints were originally caused by an organic impairment, there will be both exaggeration and a strong functional basis to symptoms.
- *Scale 3 higher than 1*. Certain degree of optimism; complaints will most likely be to the trunk of the body (e.g., gastrointestinal disorders, diseases of the lungs or heart); strong use of denial and repression, passive, sociable, and dependent; manipulate others through complaints about their “medical” problems.
- *Scale 3 is lower than 1*. Significantly more negative view of the world; any conversion is likely to be to the body extremities, such as the hands or legs.
- *High 13/31 and very high 8 and 2*. Somatic delusions.
- Symptom-related complaints usually increase with stress. When stress decreases, symptoms often disappear.

- *Frequent diagnoses.* Affective disorders (major depression, dysthymic disorder), hypochondriasis, conversion disorder, passive-aggressive personality, histrionic personality.
- High 13/31 occurs among pain patients with organic injuries whose symptoms typically worsen under stress.
- Malingering of somatic complaints might be indicated if potential gain is a factor and 13/31 is quite high (especially if 3 is above $T = 80$), even if F is not elevated (because patients want to emphasize their psychological normality but exaggerate the specifically physical nature of their difficulties).

Personality and Interpersonal Characteristics

- Superficial relationships, with extensive repression of hostility.
- Interactions often have an exhibitionistic flavor.
- Others describe them as selfish, immature, and egocentric but also as being outgoing, extraverted, and with strong needs for affection.
- Lack insight into their problems, use denial, and often blame others for their difficulties (check the Repression/ R scale).
- Usually extremely threatened by any hint that they are unconventional. Will organize themselves around ideals of service to others. Relationships and actual degree of involvement tend to be superficial.
- May feel resentment and hostility toward persons whom they feel have not provided them with sufficient attention and emotional support.
- *Conversion V in normal range* (1 and 3 at or slightly below 65 on the MMPI-2). Optimistic but somewhat immature and tangential, responsible, helpful, normal, sympathetic.

Treatment Implications

- *Difficulties in psychotherapy.* Lack insight; avoid introspection; need to appear hypernormal; prefer simple, concrete answers to their difficulties.
- Might respond to either direct suggestions or placebos, especially if the placebos are given in a medical context.
- Stress inoculation to reduce stress might be helpful.
- Describe any psychosocial interventions using medical terminology (e.g., biofeedback or other stress reduction techniques might be referred to as *neurological retraining*).
- Treatment is often terminated prematurely, especially if their defenses are challenged.
- Psychotherapeutic challenges are further increased with presence of personality disorder (check for comorbid narcissistic, dependent, antisocial, or borderline personality).

14/41 (Rare Code)**Symptoms and Behaviors**

- Severely hypochondriacal; egocentric, demand attention, express continuous concern with physical complaints.
- May have history of alcohol abuse, drug addiction, and poor work and personal relationships (check WRK/Work Interference, MAC-R/MacAndrew Alcoholism scale, APS/Addiction Potential scale, and AAS/Addiction Acknowledgment scale to refine interpretations).
- Indecisive, rebellious.
- *Frequent diagnoses.* Hypochondriasis, personality disorder (especially antisocial personality); relatively higher 4 suggests antisocial personality; relatively higher 1 suggests hypochondriasis. Profiles involving “neurotic” features (anxiety, somatoform, dissociative, and dysthymic disorders) will have relatively higher Scale 1 with 2 and/or 3.

Personality and Interpersonal Characteristics

- *Ongoing personality difficulties.* Acting out, poor judgment, extremely manipulative (but rarely extremely antisocial).
- Resentful of any rules and limits that are imposed on them.
- Rebellious toward their homes and parents, but these feelings are not likely to be expressed openly.
- Able to maintain control of impulses but will do so in a bitter, pessimistic, self-pitying manner.
- Described by others as demanding, grouchy, and dissatisfied (check the CYN/Cynicism, ASP/Antisocial Practices, and FAM/Family Problems scales).

Treatment Implications

- Resistant to therapy but may have satisfactory response to short-term, symptom-oriented treatment.
- Long-term therapy will be difficult, characterized by sporadic participation.
- Sessions may become somewhat tense because of their level of resentment and hostility, which is sometimes expressed toward the therapist (check the TRT/Negative Treatment Indicators and ANG/Anger scales).

18/81**Symptoms and Behaviors**

- Variety of vague, unusual complaints (check the HEA/Health Concerns Scale).
- Confused, disoriented, distracted, difficulty concentrating.
- Focus on physical symptoms serves as a way to organize thoughts, although the beliefs related to these symptoms may represent delusions.

- Ability to deal with stress and anxiety is extremely limited.
- Interpersonal relationships experienced with considerable distance and alienation.
- Often feel hostile and aggressive but usually keep these feelings inside. When they do express their feelings, it is in an extremely inappropriate, abrasive, belligerent manner.
- Perceived by others as eccentric or even bizarre.
- Will distrust others and may disrupt relationships because of their difficulty controlling hostility.
- Possible paranoid ideation (possibly, although not necessarily, reflected in an elevated Scale 6).
- *Elevated 18/81 along with 2.* Emphasize self-critical, pessimistic dimensions.
- *Elevated 18/81 along with 7.* Emphasize the presence of fears and anxiety (check the ANX/Anxiety, A/Anxiety, FRS/Fears, and OBS/Obsessions scales).
- *Elevated 18/81 along with 3.* Conversions and/or somatic delusions.
- *Frequent diagnoses.* Schizophrenia (especially with elevated *F*), hypochondriasis (with lower *F*), anxiety disorder (with elevated 7).

Personality and Interpersonal Characteristics

- Personality difficulties of a long-standing nature.
- Low trust, feelings of social inadequacy.
- Feel socially isolated and alienated.
- Histories often reveal a nomadic lifestyle, poor work histories (check the WRK/Work Interference scale).

Treatment Implications

- Difficulty engaging them in therapy due to poor insight.
- Distrustful, pessimistic, alienated, and even hostile (check the TRT/Negative Treatment Indicators scale).

19/91 (Rare Code)

Symptoms and Behaviors

- Possible organic difficulties relating to endocrine dysfunction or the central nervous system; gastrointestinal difficulties, exhaustion, headaches.
- Extensive complaining, overconcern with physical difficulties.
- May paradoxically attempt to deny and conceal complaints; may invest significant energy in avoiding confrontations relating to their complaints yet will make a display of these techniques of avoidance.
- Extraverted, talkative, and outgoing but also tense and restless.
- Possibly in a state of turmoil and experiencing anxiety and distress.
- Extremely high expectations yet goals will be poorly defined and often unobtainable.

- If complaints have no organic basis, their behavior may be an attempt to stave off an impending depression. Depression will be related to strong but unacceptable dependency needs.
- *Frequent diagnoses.* Hypochondriasis, manic states (may occur simultaneously). May be in response to, and exacerbated by, an underlying organic condition, an impending depression, or both; passive-aggressive personality (especially if 4 and 6 are elevated).

Personality and Interpersonal Characteristics

- Superficial appearance of being outgoing, assertive, and ambitious.
- Underlying passive-dependent core.

Treatment Implications

- Reluctant to accept psychological explanation for complaints (check the TRT/Negative Treatment Indicators scale).

23/32

Symptoms and Behaviors

- Lacking in energy, weak, apathetic, listless, depressed, anxious.
- Frequently report gastrointestinal complaints.
- Feel inadequate, difficulty accomplishing daily activities.
- Much of their energy is invested in excessively controlling their feelings and behavior.
- Although situational stress may serve to increase their depression, usually this depression is long-standing; they have learned to live with their unhappiness and general lack of satisfaction.
- Not very involved or interested in life and have difficulty initiating activities.
- *Males.* Ambitious, industrious, serious, competitive, immature dependent; strive for increased responsibilities yet also fear them; want to appear normal and receive recognition for their accomplishments yet they often feel ignored; their level of work adjustment is often inadequate.
- *Females.* More apathetic and weak, and experience significant levels of depression, resigned to long-term unhappiness and a lack of satisfaction, probably significant marital strife (check the FAM/Family Problems scale), but they rarely seek divorce.
- *With high Scale 4.* Angry, brooding component to depression, underlying antisocial thoughts, yet their external behavior is usually overcontrolled.
- *With high Scale 6.* Depression relates to extreme interpersonal sensitivity, distrust.
- *With high 0.* Socially withdrawn, introspective.
- *Frequent diagnoses.* Somatoform disorder, frequently among patients with chronic pain, affective disorders; with high *F* and/or 8: major depression with psychotic features.

Personality and Interpersonal Characteristics

- Passive, docile, dependent; typically obtain nurturance from others.
- Obtain security by keeping relationships superficial.
- Uncomfortable around members of the sex they are attracted to, may experience sexual maladjustment, including impotence or frigidity.
- Perceived as immature, childish, socially inadequate.
- Feel the need to achieve and be successful but are afraid of the added pressure this might produce.
- Superficially appear as if they are driven to succeed, but are anxious regarding competitive situations.

Treatment Implications

- Rarely volunteer for psychotherapy.
- Poor insight.
- Usually do not show significant improvement during treatment; therapy represents a threat to their use of denial and avoidance.
- Highly invested in medical explanations for their complaints.
- Seek medical “solutions” to interpersonal conflicts through methods such as tranquilizers and pain medications.
- Difficulty tolerating any discomfort but seem resigned to live with their unhappiness.
- Conflicts are likely to be somatized.
- Since distress is usually quite high, some method of symptom relief is indicated, possibly antidepressant medication.
- Supportive (rather than insight-oriented) psychotherapy is often beneficial.

24/42**Symptoms and Behaviors**

- Difficulty maintaining control over antisocial impulses.
- Experience guilt and anxiety regarding the consequences of their actions; anxiety usually occurs too late to serve as an effective deterrent; difficulty planning ahead.
- When guilt and anxiety are reduced, there is usually further acting out (24/42 code often represents an antisocial personality who has been caught).
- Depression is probably situational; distress probably reflects fear of external consequences rather than an actual internalized moral code.
- Check for heavy drinking and/or drug abuse, which serves as a form of self-medication for depression (check the MAC-R, ACK/Alcohol Acknowledgment, and APS/Alcohol Potential scales).
- Poor interpersonal relationships, numerous family difficulties (check the FAM/Family Problems scale).

- Sporadic employment; prospects for long-term employment are rarely favorable (check the WRK/Work Interference scale).
- Check for numerous legal complications (check the ASP/Antisocial Practices scale).
- *With high 6.* May feel justified in externalizing their anger because of real or imagined wrongs committed against them.
- *With low 6.* Suppression or unconscious denial of hostility.
- *With high 9.* Extremely dangerous and volatile, may have committed violent behaviors.
- *Frequent diagnoses.* Passive-aggressive or antisocial personalities (especially with high 6), adjustment disorder with a depressed mood; if depression is chronic, then anxiety, conversions, and depression (neurotic features) will be predominant (especially if Scales 1 and 3 are also high). If depression is reactive, this more likely represents an antisocial personality who has been apprehended; substance abuse may be either the primary difficulty or may occur in addition to the other disorders suggested earlier; extremely elevated 4 (above $T = 90$): a psychotic or prepsychotic process may be present, especially if F and 8 are also high.

Personality and Interpersonal Characteristics

- Initial impression of friendliness or even charm; at their best, they can appear sociable, competent, enthusiastic, outgoing.
- In a hospital setting, these patients may attempt to manipulate staff.
- Produce resentment in interpersonal relationships over the long term.
- Superficially competent and confident but feel self-conscious and an underlying sense of dissatisfaction.
- Respond to failures with pessimism, self-criticism, self-doubt.
- Often develop passive-dependent relationships to deal with feelings of self-doubt.

Treatment Implications

- Frequent pattern found in alcohol and drug treatment programs.
- Regardless of setting or reason for referral, check for substance abuse (also check the AAS/Addiction Acknowledgment scale).
- Long-standing personality difficulties that often make therapy difficult; promise to change due to guilt, which is generally authentic, but their acting out is usually resistant to change.
- *Effective therapy.* Clear limits, change in environment, warm supports, continual contact.
- External monitoring of their treatment is essential (e.g., legal or work-related); perhaps conduct treatment in a controlled environment.
- Long-term success in therapy is poor; they are likely to terminate when confronted with situational stress or when external motivators (e.g., legal) have been eliminated.

- Group interventions are likely to be more effective than individual treatment since they are highly influenced by peers.

26/62

Symptoms and Behaviors

- Extreme sensitivity to real or imagined criticism.
- Even minor criticism is brooded over and elaborated on.
- Will sometimes interpret the statements of others in a way that creates rejection, but conclusions will be based on insufficient data.
- Described by others as resentful, aggressive, hostile.
- Self-fulfilling and self-perpetuating dynamic. Often reject others first as a means of protection from perceived impending rejection by others. This results in other people avoiding them; the avoidance provides evidence that they are being rejected, which gives them a justification for feeling and expressing anger. They blame others for their difficulties, yet others have difficulty understanding the part they play in creating the interpersonal responses directed toward them.
- Long histories of difficulties with interpersonal relationships.
- Well-controlled, well-defined paranoid system with a generally adequate level of adjustment if 2, 6, and *F* are only moderately elevated.
- *Frequent diagnoses.* Dysthymic disorder; passive-aggressive personality (with elevated 4); likelihood of a psychotic or prepsychotic condition, especially paranoid schizophrenia (especially if 7, 8, and 9 are also high).

Personality and Interpersonal Characteristics

- Likely to have poor interpersonal relationships due to their hostility and hypersensitivity (check the FAM/Family Problems and CYN/Cynicism scales).
- Blaming, resentful, hostile, possibly passive-aggressive qualities, usually of a long-standing nature and difficult to alter.

Treatment Implications

- Challenge to develop and maintain rapport and trust; must continually disengage from hostility and suspiciousness (check the ANG/Anger Scale).
- Assess for possible underlying psychotic processes.

27/72

Symptoms and Behaviors

- Code is extremely common in psychiatric populations; reflects persons who are depressed, agitated, restless, and nervous.
- Scales 2 and 7 reflect the relative degree of subjective turmoil the person is experiencing, referred to as the *distress scales* (also check ANX/Anxiety, A/Anxiety, FRS/Fears, OBS/Obsessiveness scales).

- Excessive worry, often overreact to real or imagined events.
- Obsessive, experience a wide variety of phobias and fears (check the FRS/Fears scale).
- Strong and inflexible consciences, will often be extremely religious in a rigidly fundamental manner.
- Possibly slowed speech and movements.
- Insomnia.
- Feelings of social and sexual inadequacy.
- Spend a good deal of time anticipating problems before they actually occur.
- Feel vulnerable to actual or imagined threats.
- Physical complaints may include weakness, fatigue, chest pain, constipation, and dizziness (check the HEA/Health Concerns scale).
- Occurs more frequently with males 27 years or older from higher educational backgrounds.
- *With accompanying high 4.* Anxious and depressed because of poor judgment related to self-indulgence, particularly related to problem alcohol or drug use (check MAC-R, AAS/Addiction Acknowledgment, APS/Addiction Potential scales).
- *Frequent diagnoses.* Affective disorders (particularly major depressive disorder), adjustment disorder with depressed mood, anxiety disorders (particularly obsessive-compulsive disorder), personality disorders (avoidant, compulsive, passive-aggressive); may be normals who are fatigued and exhausted but who also have a high degree of rigidity and excessive worry (with moderate elevations).

Personality Characteristics

- Perfectionistic, meticulous.
- High need for recognition.
- Difficulty asserting themselves, self-blaming, self-punishing, passive-dependent (check the SOD/Social Discomfort scale).
- Will rarely be argumentative or provocative.
- Most are married, courtships were fairly brief, many marrying within 1 month of their initial dating.
- Described by others as docile and dependent, typically elicit nurturance, excessive reliance on friends and family.
- They feel inadequate, insecure, deal with feelings of hostility in an intro-punitive manner.

Treatment Implications

- Good prognosis for therapy with moderate elevations; they are introspective and experiencing a sufficient amount of distress to be motivated to change.
- Clients typically express a great deal of pessimism regarding treatment and the future in general.

- Their psychological distress is ordinarily reactive; in time, they can be expected to improve.
- The disorder usually takes between 1 month and 1 year to develop.
- Will typically be their first need for intervention.
- *Extremely high scores.* Person may be too agitated to focus and concentrate; medication may be necessary for person to function in a psychotherapeutic context.
- Suicidal thoughts are a definite possibility (especially with high 6 and 8); carefully evaluate for dangerousness.
- Are often extremely self-critical during therapeutic sessions, require considerable emotional support.
- Prone to being perfectionistic, guilty, with frequent unproductive periods of rumination.
- Although obsessive about change, they often have a difficult time actually attempting new behaviors.
- Generally establish new relationships relatively easily; these are frequently deep and of long duration.
- *With high 4.* Drinking patterns might be of a long-standing nature, therefore complicating any interventions (assess early in treatment). Do not do well in individual insight-oriented therapy. Likely to terminate prematurely. There may be an initial honeymoon effect in which changes have apparently been made. During times of stress they are likely to act out and undermine any progress. Will most likely benefit from group interventions with a focus on clear, specific goals that would include, among other things, environmental changes.

28/82

Symptoms and Behaviors

- Depression, anxiety, insomnia, fatigue, weakness.
- Mental confusion, memory impairments, difficulty concentrating.
- Agitated, tense, “jumpy.”
- Motivation to achieve is characteristically low, poor level of efficiency.
- Unoriginal, stereotyped, apathetic, indifferent.
- Excessive guilt, self-punitive.
- Fears relating to difficulty controlling emotions and impulses, including suicide.
- May cope by denying unacceptable impulses, sometimes results in dissociative periods of acting out.
- Possible delusions and hallucinations (especially if Scale 8 is greater than $T = 85$).
- Note highly diverse description of attributes only some of which may be present in any specific case (see descriptions under Scales 8 and 2). It is crucial to examine data other than Scale 8 elevation (critical items, clinical interview data; personal history; Harris-Lingoes scales; content scales particularly BIZ/Bizarre Mentation, FRS/Fears, OBS/Obsessions, LSE/Low Self-Esteem, SOD/Social Discomfort).

- *Possible diagnoses.* Major affective disorder (bipolar-depressed or major depression), schizophrenia or schizoaffective disorder, personality disorders (borderline, avoidant, obsessive-compulsive, schizoid; features likely to include liability, emotional instability, acting out).

Personality and Interpersonal Characteristics

- Resentful, unassertive, dependent, irritable.
- Suspicious, extremely sensitive to criticism.
- Interpersonal ambivalence.
- May feel withdrawn, alienated.

Treatment Implications

- Multiple problems related to expressing anger, relationship difficulties, and social withdrawal.
- Might lose control over feelings of anger, which may be directed toward therapist during times of stress.
- Likely to feel ambivalence toward relationships in general resulting in resistance to therapy; ambivalence may make it difficult to experiment with new strategies learned in therapy.
- Therapy tends to be long term.
- Therapist can provide potential point of stability in an otherwise chaotic and unpredictable life.
- Assess for suicide potential both during the initial session(s) and throughout treatment.
- May require medication during times of crises to control thoughts and feelings.

29/92

Symptoms and Behaviors

- High level of energy.
- Energy may be associated with a loss of control, or it may serve to defend against experiencing underlying depressive feelings; by speeding up activity, they can distract themselves from unpleasant depressive experiences.
- May use alcohol either to relax or to decrease depression; sporadic alcohol abuse is common.
- Anxiety and depression are likely to be present; will ruminate on feelings of worthlessness.
- Tension, restlessness.
- Somatic complaints (especially upper gastrointestinal).
- Among younger persons, 29/92 might reflect vocational crisis with a resulting loss of identity.
- Sometimes persons with brain injury have this profile; reflects their feeling of loss of control over thoughts and feelings, attempts to compensate by speeding up their level of activity.

- *Frequent diagnoses.* Mixed bipolar depression—both scales can change according to the particular phase the patient is in (state-dependent scales), cyclothymic disorder, brain injury.

Personality and Interpersonal Characteristics

- Inadequacy, worthlessness.
- Person may deny these feelings and defend against them with excessive activity.
- Typically perceived as self-absorbed and self-centered.
- High needs for achievement but may paradoxically set self up for failure.

Treatment Implications

- Alternating periods of intense activity followed by exhaustion and depression.
- Major challenge of treatment is to stabilize mood and activity swings.
- Treatment may be complicated by a long-standing history of alcohol or drug abuse.
- Carefully monitor suicide potential.
- Depression may not be immediately apparent, but careful consideration of the client's background usually reveals long-term but sporadic phases of depression.

34/43

Symptoms and Behaviors

- Immature, self-centered.
- High level of anger that they have difficulty expressing; anger will often be expressed in an indirect, passive-aggressive style.
- Continually trying to conform and please other people but still experience a considerable degree of anger; they struggle to find ways of controlling or discharging it.
- Anger stems from a sense of alienation and rejection by family members.
- Poor insight regarding their own behavior. Lack of insight even more pronounced with high 6; anger will be projected onto others.
- *Females.* More likely than males to have vague physical complaints such as headaches, blackouts, and upper-gastrointestinal complaints; still relatively free from extensive levels of anxiety; relationships will be superficial and characterized by naive expectations and a perfectionistic view of the world; gloss over and deny conflicts.
- *Frequent diagnoses.* Passive-aggressive interactional style, histrionic or borderline personalities, adjustment disorder with depressed mood (or mixed emotional features), fugue states in which aggressive and/or sexual impulses will be acted out (if both 3 and 4 are extremely elevated; $T > 85$).

Personality and Interpersonal Characteristics

- Significant conflicts relating to dependence versus independence.
- Demand approval and affection but also have underlying feelings of anger that can easily become activated by criticism.

- Superficially conforming, but underneath they have strong feelings of rebelliousness.
- Past interpersonal relationships have been difficult; history of acting out, marital discord, alcohol abuse (check the MAC-R, AAS/Addiction Acknowledgment scale, APS/Addiction Potential scale, and MDS/Marital Distress scale).
- Will project blame onto others; low insight regarding this coping style.
- Might at times vicariously act out their aggression by developing a relationship with an individual who directly and spontaneously expresses his or her hostility. Such a relationship might be characterized by the 34/43 individual's covertly encouraging and fueling the other person's angry expressions yet, on a more superficial social level, disapproving of the other person.

Treatment Implications

- Stormy treatment sessions because the therapeutic relationship will be treated in a similar way as other relationships.
- Central issues will be self-control and difficulty taking responsibility for behaviors.
- Will terminate therapy out of anger and frustration.
- Internal motivation to seek therapy is often lacking, have been forced into treatment through external pressures (e.g., spouses, work, legal justice system).
- Arrange for some external monitoring and external motivation to keep them in treatment.
- Group therapy can be quite effective since they are relatively more responsive to peer (versus authority) pressures.

36/63

Symptoms and Behaviors

- Extremely sensitive to criticism.
- Will repress hostile and aggressive feelings.
- Fearful, tense, anxious.
- May complain of physical difficulties (headaches or stomach problems).
- Overtly deny suspiciousness and competitiveness, see the world in naively accepting, positive, perfectionistic terms.
- *When 6 is higher than 3 (>5 points).* Will attempt to develop some sense of security in their lives by seeking power and prestige.
- *When 3 is higher than 6 (>5 points).* Will deny any conflicts or problems, will idealize both themselves and their world, will be more likely to develop somatic complaints rather than paranoid ideation, chance of a psychotic process is significantly reduced.

Personality and Interpersonal Characteristics

- Will harbor feelings of resentment and hostility, especially toward family members, but unlikely to express these feelings directly.

- Naive and gullible.
- Can develop comfortable, superficial relationships quickly and easily.
- As relationships' depth and closeness increase, underlying hostility, egocentricity, and even ruthlessness become more apparent.

Treatment Implications

- Limited ability to acquire personal insight, psychologically unsophisticated, resent suggestions that their difficulties may be even partly psychological (check the TRT/Negative Treatment Indicators scale).
- Will blame their personal problems on others, thereby creating difficulties in therapeutic relationship.
- Will typically terminate abruptly and unexpectedly.
- Can be ruthless, defensive, uncooperative.
- Core issue will be having them take responsibility for their feelings and behaviors.

38/83 (Rare Code)

Symptoms and Behaviors

- Anxiety, depression.
- Complaints include headaches, gastrointestinal disturbances, numbness.
- May have a series of obscure, intractable somatic complaints.
- Thought disturbances including mental confusion, disorientation, difficulties with memory and, at times, delusional thinking (especially when 8 is significantly higher than Scale 3; check BIZ/Bizarre Mentation).
- Experience turmoil; feel tense, fearful, worried.
- Outwardly apathetic and withdrawn.
- Will describe their difficulties in a vague, guarded, and nonspecific manner.
- *With elevated K and low F.* Affiliative, inhibited, overconventional, exaggerated need to be liked and approved of by others; unrealistic yet unassailable optimism; emphasize harmony, perhaps even at the cost of sacrificing their own needs, attitudes, and beliefs; extremely uncomfortable with anger and will avoid it at all costs; will avoid independent decision making and many other situations in which they must exert their power. Due to exaggerated optimism and denial of personal conflicts, they rarely appear in mental health clinics; almost as if any feelings of anger, tension, or defeat are intolerable; these feelings seem to represent both a personal failure and, perhaps more important, a failure in their attempts at controlling their world by developing an overconventional, exaggeratedly optimistic, inhibited stance.
- *Frequent diagnoses.* Somatoform or dissociative disorders (when 3 is relatively higher than 8, and 8 and/or $F < 70$), possible schizophrenia (when 8 and F both highly elevated).

Personality and Interpersonal Characteristics

- Immature, dependent, strong needs for attention and affection.
- Superficially conventional, stereotyped, unoriginal.

- Despite unusual internal experiences, they are uncomfortable with these processes and will limit them by being intro-punitive.
- Their unusual experiences and thoughts will make them feel socially alienated, but they have strong needs to appear normal and strong needs for affection.
- If others knew how unusual their experiences were, they feel they would be rejected; thus they develop extremely dependent relationships.
- To protect themselves, they use extensive denial, which makes their capacity for insight poor.

Treatment Implications

- Difficult to engage in therapy since they are typically apathetic and uninvolved in life activities.
- Treatment further complicated by low level of insight.
- They place considerable effort into appearing normal despite considerable unusual underlying processes.
- Individual insight-oriented therapy is contraindicated.
- May be responsive to a highly supportive, directive approach.

45/54

Symptoms and Behaviors

- Difficulty incorporating societal values.
- Can usually control antisocial feelings but may have brief episodes of acting out associated with low frustration tolerance and underlying anger and resentment.
- Usual coping style is through passive-aggressive means.
- The 45/54 code should in no way be considered indicative of homosexuality (see Scale 5 descriptor).
- *Note.* Scale 5 is not considered as a “clinical” scale but more provides the “tone” for the clinical scales. For example, a profile in which 4, 5, and 6 are all high might be interpreted as if it were a 46/64 code type, but the high 5 for a male would decrease the likelihood of acting out.
- *Males.* This code type occurs much more frequently among openly nonconformist, psychologically sophisticated men; if they are from higher educational levels, they will be more likely to direct their dissatisfaction toward social causes and express organized dissent toward the mainstream culture; with high 9, they will be dissatisfied with their culture, sensitive, and aware but will also have the energy to attempt to create change.
- *Males with high 4, 9 and low 5.* High probability of sexual acting out (“Don Juan” personality), self-centered, difficulty delaying gratification, behind overt display of affection is an underlying current of hostility.
- *Females.* Openly rebelling against traditional feminine role; often rebellion is motivated by intense fear related to developing dependent relationships. Alternatively they may merely be involved in a subculture or occupation that emphasizes traditionally male-oriented activities.

Personality and Interpersonal Characteristics

- Immature, self-centered, inner-directed.
- Nonconformist, likely to openly express this nonconformity in a challenging, confrontational manner.
- May have significant problems with sexual identity, possibly experience sexual dysfunction.
- Possible ambivalence relating to strong but unrecognized dependency needs.

Treatment Implications

- Guarded and defensive about revealing themselves.
- Capable of thinking clearly, good insight.
- Rarely report for treatment because they typically are satisfied with themselves and their behavior.
- Do not usually report being emotionally distressed.
- Typical issues relate to dominance and dependence.
- Significant change is unlikely because of the chronic, ingrained nature of their personality.

46/64**Symptoms and Behaviors**

- Hostile, brooding, distrustful, irritable, immature, self-centered.
- Continually blame others for their personal faults; this prevents them from developing insight into their own feelings and behavior because they are constantly focusing on the behavior of others rather than their own.
- Note history of drug addiction or alcohol abuse (check the MAC-R, AAS/Addiction Acknowledgment, and APS/Addiction Potential scales).
- *Males with high 8s.* Psychotic, especially paranoid schizophrenia or prepsychotic; with 2 and/or 3 also elevated, the chances of a borderline condition are significantly increased. Will be angry and have significant conflicts relating to their own denied but strong needs for dependency, will rebel against authority figures, may use suicidal threats to manipulate others.
- *Females.* May be psychotic or prepsychotic but are more often passive-aggressive personalities; with high 3, they will have intense needs for affection and will be egocentric, demanding, but resentful of the demands placed on them by others.

Personality and Interpersonal Characteristics

- Passive dependency.
- Adjustment difficulties associated with hostility, anger, mistrust, and blame of others.
- Avoid deep involvement.

- Are perceived as sullen, argumentative, obnoxious, resentful of authority (check the ANG/Anger scale).
- Usually unable to form close relationships; significant levels of social maladjustment.
- Minimal self-criticism, highly defensive, argumentative (especially with high *L* and *K*).
- Highly sensitive to real or imagined criticism from others, often inferring hostility or rejection when this was not intended.
- To avoid rejection and maintain a certain level of security, they become extremely adept at manipulating others.

Treatment Implications

- Suspicious and even antagonistic toward treatment.
- Treatment typically occurs at the insistence of someone else.
- Project the blame for any difficulties onto someone else.
- Treatment plans should be concrete, clear, realistic, and described in a way that does not arouse suspicion or antagonism.
- Therapeutic relationship is difficult to establish; once established, is likely to be somewhat turbulent.
- Carefully monitor angry acting out.

47/74

Symptoms and Behaviors

- Brooding and resentful.
- Experience guilt over behavior.
- Insensitive to the feelings of others but intensely concerned with their own responses and feelings; justify insensitivity because they feel rejected or restricted by others.
- *Predictable interpersonal cycle.* Express anger with little control over their behavior, resulting in impulsive acting out (check the ASP/Antisocial Practices and ANG/Anger scales); will feel guilty over behavior, followed by a phase of excessive overcontrol accompanied by guilt, brooding, and self-pity (check the O-H/Over-Controlled Hostility scale). Frustrated by these feelings, they may then attempt to selfishly meet their needs through alcohol abuse, promiscuity, or further aggressive acting out. Cycle is usually resistant to change.
- Frequently leads to legal, work-, and home-related difficulties.
- Even though guilt and remorse are genuine (even excessive), their self-control is still inadequate and acting out continues.
- *Frequent diagnoses.* Antisocial personality, anxiety disorder, alcohol/drug abuse (check the MAC-R, AAS/Alcohol Acknowledgment, APS/Alcohol Potential

scales), miscellaneous conditions with impulsive-compulsive styles (e.g., eating disorders).

Personality and Interpersonal Characteristics

- Insecurity and ambivalence regarding dependency.
- Need frequent reassurances that they are worthy.

Treatment Implications

- Early treatment characterized by sincere remorse and need to change; as guilt diminishes, acting out again occurs (be suspicious of early “easy” gains).
- Limit-setting will be met with anxiety and resentment; often they either test the limits or completely ignore them.
- Acting out followed by guilt is a chronic pattern.
- Therapeutic attempts to decrease anxiety may actually result in increased acting out because the control created by guilt and remorse might be diminished.
- May respond well to reassurance and support.
- Long-term, fundamental change will be difficult to achieve.

48/84

Symptoms and Behaviors

- Strange, eccentric, emotionally distant, severe problems with adjustment.
- Behavior is unpredictable, erratic; may involve strange sexual obsessions and responses.
- Antisocial behavior has resulted in legal complications (check the ASP/Antisocial Practices).
- Very little empathy, nonconforming, impulsive.
- Possibly members of strange religious cults or unusual political organizations.
- *Early family histories.* Learned that relationships were dangerous; constant confrontation with intense family conflicts; felt alienated, hostile, rejected; attempted to compensate with counterrejection and other forms of retaliation.
- Erratic academic performance, characterized by underachievement.
- *With high F and low 2.* Aggressive, cold, punitive, inspire guilt and anxiety in others. Often they take on roles in which such behavior is socially sanctioned (e.g., rigid law enforcement officer, overzealous member of the clergy, strict school disciplinarian). Behavior may range all the way from merely stern, punitive, and disapproving, to actual clinical sadism. Underneath these overt behaviors they usually have a deep sense of alienation, vulnerability, and loneliness, which may give rise to feelings of anxiety and discomfort.
- *Males.* Frequent criminal behavior (especially with high 9). Crimes are often bizarre, impulsive, poorly planned, self-defeating, eventually result in self-punishment, occur without apparent reason, are extremely violent, involve homicide and/or sexual assault.

- *Females*. Less likely to act criminally. Relationships will usually be primarily sexual, rarely become emotionally close; relationships will be with partners who are significantly inferior to themselves (often described as losers).
- *Frequent diagnoses*. Schizoid or paranoid personality, psychotic reaction, paranoid schizophrenia (especially with high 6).

Personality and Interpersonal Characteristics

- Deep needs for attention and affection.
- Frequently set themselves up for rejection and failure.
- Deep feelings of insecurity, poor self-concept.
- Poor interpersonal judgment, inadequate communication; others feel as if they are missing important elements or significant connotations of what the 48/84 individual is saying, but they cannot figure out exactly what or why.

Treatment Implications

- Difficult to establish therapeutic relationship since clients are aloof and unconventional.
- Sessions are likely to be chaotic with difficulty focusing on relevant areas. There will be so many different problems to work on it will be difficult to know where to begin. It is easy to get sidetracked; as a result, sessions may seem relatively unproductive.
- Long-standing drug- and alcohol-related problems may complicate treatment.
- Acting out may further complicate treatment.
- Since these clients are mistrustful, they are likely to terminate prematurely.

49/94

Symptoms and Behaviors

- Feel alienated.
- Antisocial tendencies with the energy to act on these tendencies.
- Self-indulgent, sensation seeking, impulsive, oriented toward pleasure, irritable, extraverted, violent, manipulative, energetic.
- Poorly developed conscience, marked lack of concern for rules and conventions.
- Free from anxiety, talkative, articulate, charming; can often make a good initial impression.
- Relationships are usually shallow; any sort of deeper contact with them brings out the more problematic sides of their personality.
- History typically reveals extensive legal, family, and work-related difficulties (check ASP/Antisocial Practices and WRK/Work Interference).
- Pattern is highly resistant to change when found in persons over age 30.
- *Adolescent males*. Associated with delinquency.
- *With low 0*. Highly developed social techniques, will use these skills to manipulate others; may be involved in elaborate, antisocial con games.

- *With high 3.* Decreases the chance of acting out; expression of hostility is likely to be similar to that of the 34/43 code in that it will be indirect and often passive-aggressive.
- *With high 6.* Therapists should use extreme caution. Clients are very dangerous, with poor judgment. Acting out will often be violent and bizarre, will appear justified to themselves because of strong feelings of resentment.
- *Frequent diagnoses.* Antisocial personality (but use caution when categorizing adolescents, as these scales are commonly elevated for both normal and abnormal adolescents), manic state, or schizophrenia (with high 8).

Personality and Interpersonal Characteristics

- External facade of confidence and security but underlying immaturity, dependence, and insecurity.
- Narcissistic, great difficulty establishing deep emotional closeness.
- Difficulty delaying gratification, will often exercise poor judgment.
- Perceived by others as extraverted, talkative, uninhibited, restless, needing emotional stimulation and excitement.
- Initial good impression, but their antisocial style soon becomes apparent.
- Will rationalize their shortcomings and blame their problems on others.

Treatment Implications

- *Numerous difficulties encountered in therapy.* Problem focusing, constantly embarking on irrelevant tangents, difficulty delaying gratification, do not learn from experience, primarily concerned with self-gratification (often at the expense of others), frequently irritable, if confronted by a therapist will express their fairly extensive hostility, typically cope through conning other people.
- May use charm laced with occasional belligerence. When this behavior occurs, it is advisable to confront it as soon as possible.
- Treatment is likely to be slow, frustrating, often unproductive.
- Rarely volunteer for therapy, typically referred by the court system or at the insistence of someone else (e.g., employer, spouse).
- External monitoring is usually required to keep them in treatment.
- Because their anxiety level is quite low, they will not be motivated to change.
- Group treatment has been reported to be relatively helpful; behavioral modification can often help them develop better coping styles.
- Termination is usually premature and associated with clients feeling bored with the sessions, acting out, or a combination of the two.

68/86

Symptoms and Behaviors

- Suspicious, distrustful; perceive the intentions of others as suspect and questionable.
- Interpersonally distant, few or no friends.

- Inhibited, shy, resentful, anxious, difficulty accepting or appropriately responding to the demands made of them.
- Highly involved in fantasy world.
- Uncooperative, apathetic, poor judgment.
- Experience difficulty concentrating.
- Sense of reality is poor.
- Often experience guilt, inferiority, mental confusion.
- Flat affect.
- Unusual, even bizarre thoughts, delusions of grandeur and/or self-reference.
- Internally quite anxious.
- Past work history is often (surprisingly) adequate (when the elevations on 6 and 8 are not extremely high).
- Intensification of their symptoms brought on by stress usually disrupts their ability to work.
- Typically single and younger than 26 years of age.
- If they are married, their spouses are frequently also emotionally disturbed.
- Highly elevated *F* with Scales 6 and 8 above $T = 80$ does not necessarily indicate an invalid profile.
- *With elevated 2.* Depression with inappropriate affect, phobias, and paranoid delusions.
- *Frequent diagnoses.* Paranoid schizophrenia (especially with high 4 and if 8 is relatively higher than 7); note “paranoid valley” (when 7 is 10 points or more < Scales 6 and 8), which emphasizes the presence of paranoid ideation; possibly organic brain disorders.

Personality and Interpersonal Characteristics

- Insecure, low self-confidence, poor self-esteem (check LSE/Low Self-Esteem).
- Others perceive them as being unfriendly, negativistic, moody, and irritable.
- High social discomfort; will feel most relaxed when alone, will generally avoid deep emotional ties (check SOD/Social Discomfort).
- Poorly developed defenses.
- Usually regress under stress.

Treatment Implications

- *Numerous issues related to further assessment and case management due to significant level of psychopathology.* Inpatient or outpatient treatment, danger to self or others, possible psychopharmacological intervention and maintenance, basic daily living skills.
- Training in basic social skills, assertiveness, job interviewing, and knowledge of resources to resort to when their symptoms increase.
- Insight-oriented therapy is often contraindicated as self-reflection might result in further regression.

- Concrete, behaviorally oriented methods of intervention are likely to be more successful.
- May have unusual or even bizarre belief systems with quite different sets of logic from the therapist (check the BIZ/Bizarre Mentation scale), which is likely to make cognitively based interventions difficult.
- Level of suspicion and projection of blame will present further challenges.
- Mistrust, poor social skills, and social discomfort may make it difficult to form a therapeutic relationship.
- Sessions often seem slow and unproductive and characterized by long periods of silence.
- Impulsivity and regression likely to provide further treatment challenges.

69/96

Symptoms and Behaviors

- Excited, oversensitive, mistrustful, energetic, irritable.
- Difficulty thinking.
- Obsessional, ruminative, overrideational.
- May have clear or subtle signs of a thought disorder including delusions, difficulty concentrating, hallucinations, tangential associations, incoherent speech.
- May appear perplexed and disoriented.
- Feel extremely vulnerable to real or imagined threats, experience anxiety much of the time.
- Reactions to stress can result in their becoming either overly excited or apathetic and withdrawn.
- Typical response to stress is to withdraw into fantasy.
- Difficulty modulating their expression of emotions.
- Exercise poor judgment.
- *Frequent diagnoses.* Schizophrenia (paranoid type), mood disorder.

Personality and Interpersonal Characteristics

- Mistrustful and suspicious.
- High needs for affection; relationships will often be passive-dependent.
- Clear discrepancy between how they describe themselves and how others perceive them; they describe themselves as calm, easygoing, happy, in good health; others describe them as hostile, angry, and overreactive to even minor stress.

Treatment Implications

- 69/96 is characteristic of inpatient populations.
- Psychopharmacological interventions to help control disorganized thinking or regulate mood can often be extremely effective.
- Due to disorganized, regressive, and ruminative thought processes, insight-oriented therapy is usually not effective.

- Lack of trust and suspiciousness often makes it difficult to form a therapeutic relationship.
- If a trusting relationship can be developed, concrete, problem-focused approaches are most effective.

78/87

Symptoms and Behaviors

- Agitation is often sufficiently intense to disrupt their daily activities.
- Profile represents a reaction to a specific crisis; may have been previously functioning at an adequate level until some event or series of events triggered a collapse in their defenses (“nervous breakdown”).
- Low level of self-confidence; common feelings include guilt, inferiority, confusion, worry, fear.
- Insomnia, hallucinations, delusions.
- Note extent and relation between elevations on 7 and 8 since this is important diagnostically and prognostically; if 7 is higher (at least 5–10 points) than 8, then condition is more susceptible to improvement and tends to be more benign (regardless of the elevation of 8, as long as 7 maintains its relatively higher position). When 7 is higher, this suggests that the person is still actively fighting the problem and has some defenses still working; this suggests an anxiety disorder rather than psychosis, since ingrained bizarre thought patterns and withdrawn behavior have not yet become established. A relatively higher Scale 8 (at least 5–10 points) reflects a more fixed pattern, which is more difficult to treat (particularly if Scale 8 > 75). If 7 and 8 are both greater than 75 (with Scale 8 relatively higher), this suggests an established schizophrenic pattern (especially if the neurotic triad is low; check BIZ/Bizarre Mentation). Even if schizophrenia can be ruled out, the condition tends to be extremely resistant to change (e.g., a severe, alienated personality disorder).
- *With elevated 2.* Dysthymic or obsessive-compulsive disorder.

Personality and Interpersonal Characteristics

- Feel inferior, inadequate, indecisive, insecure.
- Often passive-dependent relationships, difficulties asserting themselves in intimate relationships.
- Difficulty developing and sustaining relationships, may have difficulties related to sexual performance.
- Preoccupied with excessive and unusual sexual fantasies.
- Often feel extremely uncomfortable in most social relationships (check SOD/Social Discomfort scale).
- Likely to defend themselves with excessive withdrawal.

Treatment Implications

- 78/87 often occurs among psychiatric patients.

- Possible significant suicidal risk; check elevation on 2, relevant critical items, take a careful history, ask relevant questions related to the client's thought processes.

89/98**Symptoms and Behaviors**

- Highly energetic, perhaps to the point of hyperactivity.
- Emotionally labile, tense, disorganized.
- Possible delusions of grandeur, sometimes with a religious flavor (especially with a high 6).
- Tangential, bizarre speech, possibly characterized by neologisms, clang associations, and echolalia (check BIZ/Bizarre Mentation).
- Unrealistic goals and expectations often lead them to make extensive plans that are far beyond their ability to accomplish; aspirations will be significantly higher than their actual achievements.
- Severe symptoms related to insomnia.
- High likelihood of serious psychopathology.
- *Frequent diagnoses.* Schizophrenia, schizoaffective disorder with manic states, Bipolar, severe personality disorder; the relative elevation of *F* can be used as an index of severity.

Personality and Interpersonal Characteristics

- Childish, immature interpersonal relationships.
- Fearful, distrustful, irritable, distractible.
- Highly talkative and energetic but will also prefer to withdraw from interpersonal relationships, resist any deep involvement.
- Grandiose and boastful but with underlying feelings of inferiority and inadequacy.
- Demand considerable attention, will become hostile and resentful when needs are not met (check ANG/Anger).

Treatment Implications

- Because they are highly distractible and tangential, psychotherapeutic approaches with them are extremely difficult.
- Poor insight, will resist psychological interpretations, cannot focus on any one area for any length of time.
- Defend themselves using denial, grandiose thoughts, and an inflated sense of self-worth; challenging these defenses is likely to provoke irritability, anger, or even aggression.
- If extensive delusions and hallucinations are present, antipsychotic medication may be indicated, or mood stabilizers if a mood disorder is predominant.

MMPI-2 CONTENT SCALES

One of the earliest efforts to develop a series of MMPI content scales was by Wiggins (1966, 1971), who organized scales based on an overall analysis of the contents of the

MMPI items. He began with item clusters that were based on areas such as authority conflicts and social maladjustment. These clusters were revised and refined using factor analysis and evaluations of internal consistency. During the 1989 restandardization of the MMPI, many of the items relating to the Wiggins scales were altered or deleted. As a result, Butcher et al. (1990) developed a new set of 15 different content scales. At first, provisional content scales were developed by rationally sorting the items into different content categories. These categories were then refined statistically using item-scale correlations with psychiatric inpatients and correlations between the scales. Further validity studies have confirmed that they are at least as valid as the MMPI/MMPI-2/MMPI-A empirically derived clinical scales (Barthlow et al., 1999; Ben-Porath, Butcher, & Graham, 1991; Ben-Porath et al., 1993; Butcher & Williams, 2000). A further advantage over the clinical scales is that they measure single dimensions. The practical significance is that they can be interpreted relatively easily using rational, intuitive strategies. In contrast, the MMPI clinical and validity scales are multidimensional. Thus, they require clinicians to work with them to extract the most useful and valid interpretations, often from a wide variety of possible descriptors.

An important function of the content scales is the ability to use them to refine the meanings of the clinical scales. For example, if an individual obtains an elevation on 4 (Psychopathic Deviate), clinicians can note possible corresponding elevations on FAM (Family Problems) and ASP (Antisocial Practices). If FAM is elevated but not ASP, the elevated 4 has more to do with family alienation and conflict than criminal and other forms of antisocial behavior. Thus, the content scales can incrementally increase the validity of the clinical scales (Barthlow et al., 1999; Ben-Porath et al., 1993).

In addition to clarifying the meanings of the scales, their interpretations and implications can also be extended. For example, elevations on 1, 2, and 3 are consistent with pain patients. However, in considering their prognosis for rehabilitation programs, it would also be important to assess pain patients' attitudes toward returning to work by noting the scores on WRK (Work Interference) and responsiveness to treatment by noting scores on TRT (Negative Treatment Indicators; M. Clark, 1996; Deardorff, 2000). Elevations 65 and above on the content scales indicate that many of the descriptors for the scale apply to the person. Scales that are mildly elevated (60–64, inclusive) suggest that several of the behaviors apply to the person. Thus, the inclusion of the MMPI-2 and MMPI-A content scales represents potentially important and easily interpreted dimensions of assessment. The content scales can be divided into the clusters described next relating to internal symptoms, external aggression, negative self-views, and general problem areas.

Internal Symptomatic Behaviors

ANX/Anxiety. Generalized anxiety, somatic difficulties, worries, insomnia, ambivalence, tension, a feeling that life is a strain, fear of losing his or her mind, pounding heart and shortness of breath, concentration problems, difficulties making decisions; symptoms clearly perceived and admitted to by the client.

FRS/Fears. Multiple specific fears (nuisance animals, blood, dirt, leaving home, natural disasters, mice, snakes, etc.).

OBS/Obsessiveness. Ruminates, difficulty with decision making, resistant to change, needless repetitive counting, may have compulsive behaviors such as counting or

alphabetizing his or her experience; worried, sometimes overwhelmed by his or her own thoughts; others become easily impatient with the person. Persons with low scores are likely to be relaxed, secure, and unlikely to be depressed.

DEP/Depression. High number of depressive thoughts, uninterested in life; feeling of emptiness; feeling of having committed unpardonable sins; cries easily; unhappy; possible suicidal ideation; sense that other people are not sufficiently supportive; sensitive to rejection, tense, passive feeling of hopelessness; helplessness about the future.

HEA/Health Concerns. Numerous physical complaints regarding gastrointestinal, neurological, sensory, skin, cardiovascular, and/or respiratory difficulties; problems of adjustment; worried and nervous; lacking in energy.

BIZ/Bizarre Mentation. Psychotic thought processes, hallucinations (auditory, visual, olfactory), paranoid beliefs, strange thoughts, delusions.

External Aggressive Tendencies

ANG/Anger. Difficulties in controlling anger, irritable, impatient, annoyed, stubborn, may swear; episodes of loss of control, possibly breaking objects or actually being physically abusive. Persons scoring low are unlikely to be depressed or have significant family problems.

CYN/Cynicism. Distrust of other people; fear of being used or that others will lie to and cheat them; belief that the only reason for others not lying or cheating is fear of being caught; negativity toward friends and associates, belief that people are friendly only for selfish reasons. Persons with low scores might be highly achievement oriented.

ASP/Antisocial Practices. Past legal and/or academic problem behaviors; expectation that others will lie; support of illegal behavior; enjoyment of criminal behavior of others; thought patterns that characterize criminal behavior, whether such behavior actually occurs or not. ASP has been found to be a better predictor (greater sensitivity and specificity) of antisocial personality disorder than *Pd* (Psychopathic Deviate; S. R. Smith, Hilsenroth, Castlebury, & Durham, 1999) with a recommended cutoff of 55 or 60 (rather than the suggested cutoff of 65 implied by the MMPI-2).

TPA/Type A. Driven, hardworking, competitive, hostile, irritable with time constraints, overbearing, annoyed with interruptions, tries to do more and more in less and less time, blunt and direct, petty regarding minor details. This scale is a better construct for use with males than females.

Negative Self-View

LSE/Low Self-Esteem. Low self-confidence, feeling of insignificance, negative beliefs regarding self (clumsy, inept, unattractive), acutely aware of faults, feeling of being disliked by others, sometimes overwhelmed by his or her own faults, difficulty accepting compliments from others. Conversely, low scores suggest the person is secure, relaxed, and unlikely to be depressed.

General Problem Areas Cluster

SOD/Social Discomfort. Shy, withdrawn, uneasy with others, introverted, dislikes social events, prefers to be alone. Persons with low scores are likely to be secure,

relaxed, achievement oriented, assertive, and unlikely to be depressed or experience somatic symptoms.

FAM/Family Problems. Family discord, unhappy childhood, difficult and unhappy marriages, families that do not express much love but are rather quarrelsome and unpleasant, possibly an abusive childhood.

WRK/Work Interference. Personal difficulties that interfere with work; tension, worry, obsessiveness, difficulty concentrating, career indecision and/or dissatisfaction, poor concentration, dislike of coworkers; difficulty initiating work-related activities; little family support for career choice; easily defeated by difficulties.

TRT/Negative Treatment Indicators. Dislike or distrust of helping professionals, discomfort in discussing difficulties, low level of disclosure, resistance to change, disbelief in the possibility of change, belief that no one can really understand or help him or her, preference for giving up rather than facing a crisis.

MMPI-A CONTENT SCALES

The MMPI-A content scales were developed and refined in much the same way as the MMPI-2 content scales. Some of the items were changed to be more relevant for adolescent populations. In addition, some new scales, such as the Adolescent-School Problems scale (instead of the adult WRK/Work Interference scale), were added, and others, such as the TPA (Type A) scale, were dropped because they were not considered relevant for adolescents. Elevations 65 and above indicate that there has been extensive endorsement of the problems indicated in the scales, whereas a mild elevation (60–64, inclusive) suggests that several of the descriptors apply to the person.

A-anx/Adolescent-Anxiety. High scores suggest tension, nervousness, worry, sleep-related difficulties (nightmares, difficulty with sleep onset, early-morning awakening); life feels like a strain; problems seem insurmountable; there are feelings of impending doom, fears of losing his or her mind, confusion and difficulty concentrating, increase in family discord; girls in clinical settings report feeling depressed and have somatic complaints.

A-obs/Adolescent-Obsessiveness. High scores suggest excessive worry, ruminations, obsessive counting of objects, extreme fear regarding making changes, difficulty making decisions, obsessing over past events or behaviors; others lose patience with them; boys in clinical settings are described as anxious, overly concerned with the future, dependent, worried, preoccupied, resentful, feel as if they deserve punishment; girls in clinical settings may have suicidal ideation and/or have actually made suicidal gestures.

A-depl/Adolescent-Depression. High scores suggest fatigue, crying spells, self-criticism, feelings of being condemned and unworthy, feelings of hopelessness; life is uninteresting, suicidal ideation may be present; there is difficulty initiating activities, dissatisfaction; boys in clinical settings should be further assessed for a possible history of abuse; girls in clinical settings have depression and low self-esteem; girls in school settings are likely to have poor grades, are unlikely to have noteworthy personal achievements, and are likely to be concerned about being overweight.

A-heal/Adolescent-Health. Elevations indicate the presence of health problems that result in school absence and limit their physical activities; complaints cover several

different physical areas, including gastrointestinal (nausea, vomiting, constipation, stomach trouble), sensory problems (poor eyesight, hearing difficulty), neurological complaints (convulsions, paralysis, numbness, dizzy spells, fainting), cardiovascular problems (heart or chest pains), skin disorders, respiratory problems, excessive worry over health, and belief that all related problems would be fine if their health difficulties could be solved; in clinical settings, these adolescents are likely to report being afraid of school; in school settings, they are likely to have academic and behavioral difficulties (e.g., school suspensions, course failures, low grades); girls in clinical settings are likely to report an increase in disagreements with parents; boys in clinical settings are described as anxious, worried, guilt prone, accident prone, perfectionistic, clinging, fearful, and more likely to have lost weight.

A-ain/Adolescent-Alienation. High scores indicate a high level of emotional distance, a feeling that no one really understands or cares for them, a sense that they are getting a raw deal from life, difficulty getting along with others, not being liked, others are unkind and even out to get them; there is a belief that others have more fun than they do; low self-disclosure is likely; others interfere with their attempts to succeed; they feel anxious when talking to a group and are likely to have poor grades in school; girls may have a problem with weight gain; girls in clinical settings have few or no friends, increase in disagreements with parents; boys in clinical settings have low self-esteem and poor social skills.

A-biz/Adolescent-Bizarre Mentation. High scores indicate very strange thoughts and experiences; possibly auditory, olfactory, and visual hallucinations; paranoid thoughts (e.g., being plotted against, someone is trying to kill them); possible beliefs that evil spirits or ghosts are trying to control them; girls in clinical settings probably come from dysfunctional families, parents and/or siblings might have arrest records; boys in clinical settings are likely to have been under the supervision of a child protective worker, likely to exhibit bizarre and possibly psychotic behavior; individuals in school settings are likely to have numerous difficulties including poor grades, suspensions, and course failures.

A-ang/Adolescent-Anger. High scores indicate that the adolescent finds it difficult controlling anger, feels like breaking or smashing things, sometimes yells to make a point, and throws tantrums to get his or her way; feels like getting into fistfights; shows irritability when others try to hurry him or her, impatient; especially likely to get into fights when drinking; likely to act out in school and/or home; adolescents in clinical settings are extremely interested in violence and aggression, histories of assault; described as angry, resentful, impulsive, moody, externalize behaviors; boys in clinical settings are described as attention seeking, resentful, anxious, self-condemning, but also dependent and clinging, may have a history of sexual abuse; girls in clinical settings are likely to be aggressive, delinquent, have been arrested, act out sexually (promiscuity), are flirtatious, wear provocative clothes, need to be supervised around potential sexual partners.

A-con/Adolescent-Conduct Problems. Elevations suggest that the client is oppositional, has legal problems, peer group is often in trouble; behavior problems including lying, stealing, shoplifting, swearing, vandalism; likely to enjoy other people's criminal behavior, might also enjoy making other people afraid of him or her; uses drugs and alcohol, has record of poor academic performance and school-related behavior problems (e.g., course failures, suspensions, lying and cheating), disobedient, impulsive;

clinical girls are described as impulsive, angry, unpredictable, sexually active, provocative, resentful, impatient, require supervision around potential sexual partners, unlikely to be depressed.

A-cyn/Adolescent-Cynicism. Persons scoring high are endorsing statements that they distrust other people. They believe that if other people are nice, it is only because they are trying to take unfair advantage of the people they are being nice to. Accordingly, high scorers feel guarded and misunderstood. Because they feel that others are out to get them and mainly concerned with self-interest, persons scoring high feel justified in having misanthropic attitudes. They may also believe that others are jealous of them.

A-lsel/Adolescent-Low Self-Esteem. High scores indicate that the individual feels unattractive, useless; that he or she has little ability, many faults, low self-confidence, is unable to do anything particularly well including planning own future, is confused and forgetful, has difficulty accepting compliments, is susceptible to social pressure, passive; high-scoring boys should be further assessed for the possibility of sexual abuse; girls are likely to report weight gain, poor grades, and no noteworthy personal achievements; boys in clinical settings are described as having poor social skills; girls in clinical settings will be depressed, are likely to have learning disabilities, have increasing numbers of conflicts with their parents, suicidal thoughts, and possibly suicidal gestures.

A-las/Adolescent-Low Aspirations. High scores indicate a low level of interest, especially academically; the adolescent dislikes studying, reading, listening to lectures (especially science); has problems initiating activities, gives up easily, dislikes facing difficult situations; has low expectations for achievement and little interest in continuing on to college; described by others as lazy, has poor grades, little interest in school activities; girls in clinical settings are likely to report sexual acting out, very unlikely to report having won a prize or award; boys in clinical settings are likely to have been truant in school and run away from home.

A-sod/Adolescent-Social Discomfort. High scores indicate that the adolescent is shy, prefers to be alone, has difficulty making friends, is extremely uncomfortable when addressing a group, dislikes parties and crowds, is difficult to get to know, is uncomfortable meeting new people, dislikes initiating conversations, might actively avoid others, is unlikely to report using drugs or alcohol; boys are likely to avoid school activities; girls in clinical settings are unlikely to be involved in acting out, are uninterested in dating, have few friends, may be depressed, have eating difficulties; may be fearful, withdrawn, physically weak; and are not likely to be involved with drugs, alcohol, or irresponsible behavior.

A-fam/Adolescent-Family Problems. High scorers are likely to have extensive difficulties with parents and other family members that include fault-finding, jealousy, little love, serious arguments, poor communication; they long for the day when they can finally leave home, feel that parents punish them unfairly; they show little acceptance of responsibility around home, feel that they cannot depend on their family in times of need; beatings and runaways are possible, however, problems usually do not extend into the legal justice system; there may be some school-related difficulties (low grades, suspensions); may reflect marital difficulties of parents; girls in school settings report possible exam failure and/or weight gain; in clinical settings, there may be more externalizing behaviors including lying, cheating, stealing, and somatic complaints,

crying, guilt, timidity, and withdrawal; boys in clinical settings are described as sad, secretive, uncommunicative, disliked, self-conscious, unloved, dependent, resentful, attention-seeking, and self-blaming; girls in clinical settings are typically described as immature, likely to fight, cruel, destructive, secretive, self-conscious, hyperactive, provocative, sexually acting out (promiscuity), and preoccupied with sex; further assessment should include possible sexual abuse for girls and possible physical abuse for boys.

A-sch/Adolescent-School. High scores indicate a wide number of school-related difficulties including low grades, truancy, easily upset by school events, learning disabilities, low level of social competence, boredom, suspensions, dislike of school, disciplinary actions, difficulty concentrating, probations, and negative attitudes toward teachers; feels that school is a waste of time; often school-related difficulties are specific to school itself and do not spill over into other areas; boys from clinical populations are likely to have run away, been irresponsible, and have a history of drug use, particularly amphetamines; they should be further evaluated for the possibility of sexual abuse; girls from clinical populations may have learning disabilities and/or academic underachievement.

A-trt/Adolescent-Negative Treatment Indicators. High scores indicate negative attitudes and feelings toward health care professionals; they do not like to share personal information with others; they feel that they can never really be understood and that others do not really care what happens to them; they will have anxiety related to people asking them personal questions; they have difficulty planning for the future and are unwilling to take responsibility for the negative things in their lives; they feel that they have many secrets they need to keep to themselves.

MMPI-2 HARRIS-LINGOES AND *S*/ SUBSCALES

One of the more popular developments of the MMPI has been the reorganization by Harris and Lingo (1968) of the standard scales into more homogeneous content categories. These subscales were constructed by intuitively grouping together items that seemed to reflect single traits or attitudes contained in the already existing MMPI Scales 2, 3, 4, 6, 8, and 9. Ben-Porath et al. (1989) further developed subscales similar to the Harris-Lingo subscales for Scale 0. No subscales were developed for 1 and 7 because these were considered to be relatively homogeneous in their item content. These same subscales have been carried over for use with the MMPI-A. This section discusses the subscales and provides a brief summary of the meanings associated with high scores. These summaries are derived from material by Harris and Lingo (1968) and extensions of these materials as summarized by Butcher et al. (1990), Butcher and Williams (2000), J. R. Graham (2011), Greene (2000), and Levitt and Gotts (1995). Scoring templates and profile sheets for the MMPI-2 and MMPI-A Harris-Lingo subscales are available from Pearson Assessments.

Although the Harris-Lingo subscales show high intercorrelations with their parent scales (Harris & Lingo, 1968) and relevant code types (McGrath, Powis, & Pogge, 1998), the internal consistency of the subscales is somewhat low (.04–.85; Gocka, 1965; Krishnamurthy, Archer, & Huddleston, 1995); however, many of scales with low internal consistency have very few items, and most are adequate (Gotts & Knudson, 2005;

J. R. Graham, 2011). Additionally, many of the Harris-Lingoes subscales are highly intercorrelated, due in part to the fact that no effort was made to eliminate overlapping items (Caldwell, 1988; Greene, 2011). Several initial validity studies are available (Boerger, 1975; Calvin, 1975; N. Gordon & Swart, 1973) that demonstrate the potential clinical usefulness of these subscales. The Social Introversion subscales have been found to account for 90% of the variance of the *Si* scale, and convergent and discriminant validity was demonstrated based on an analysis of spouses' ratings of each other (Ben-Porath et al., 1989). The practical importance of both sets of subscales is that they provide a useful supplement for interpreting the original scales. For example, a clinician can assess whether a person scoring high on Scale 4 (Psychopathic Deviate) achieved that elevation primarily because of family discord (*Pd* 1), authority problems (*Pd* 2), or social imperturbability (*Pd* 3). This breakdown is likely to be quite helpful in interpreting why a client received a high score that was unexpected based on the person's history. It might also be quite useful in interpreting the significance associated with moderate elevations ($T = 60-65$). A further reason to score and interpret the Harris-Lingoes scales is to understand the possible reasons for contradictory descriptions, such as might emerge if both Scales 2 and 9 were elevated. However, if the clinical scales are either in the normal range or quite high, the Harris-Lingoes scales are not particularly useful. While some suggest Harris-Lingoes and *Si* subscale elevations of $T > 64$ should be interpreted, others recommend only interpreting these subscales if $T > 70$ (Greene, 2011).

The Harris-Lingoes and *Si* subscales were generally not used for routine interpretations because they are quite time consuming to hand-score; however, these are now included in the Extended Report provided by Pearson. Rather than interpreting all the Harris-Lingoes and *Si* subscales, clinicians can select only those that are relevant for refining and clarifying the meanings of clinical scales that are in question. Despite some validity efforts, the amount of research available is still inadequate, and, in many cases, the internal consistency of the subscales is insufficient. Thus, any interpretations should be made cautiously and be considered as hypotheses in need of further support. This is particularly true for the MMPI-A, in which there has been even less investigation using the Harris-Lingoes and *Si* subscales than for the MMPI-2. Furthermore, item deletions and alterations between the MMPI-2 and MMPI-A, primarily for the *Si* scale, bring into question the transferability of the Harris-Lingoes and *Si* scales with the adolescent version of the MMPI.

Scale 2. Depression

D1/Subjective Depression. Unhappy, low energy, sense of inferiority, low self-confidence, socially uneasy, few interests.

D2/Psychomotor Retardation. Low energy, immobilized, socially withdrawn, listless.

D3/Physical Malfunctioning. Reports wide variety of physical symptoms, preoccupied with health, denial of good health.

D4/Mental Dullness. Low energy, pessimistic, little enjoyment of life; difficulties with concentration, attention, and memory; apathetic.

D5/Brooding. May feel as if he or she is losing control of his or her thoughts; broods, cries, ruminates, feels inferior, and is hypersensitive.

Scale 3. Hysteria

Hy1/Denial of Social Anxiety. Extraverted, comfortable with social interaction, minimally influenced by social standards.

Hy2/Need for Affection. Strong needs for affection with fears that these needs will not be met, denies negative feelings toward others.

Hy3/Lassitude-Malaise. Subjective, discomfort, poor health, fatigued, poor concentration, insomnia, unhappiness.

Hy4/Somatic Complaints. Wide variety of physical complaints, denial of hostility toward others.

Hy5/Inhibition of Aggression. Denial of hostility and anger, interpersonally hypersensitive.

Scale 4. Psychopathic Deviate

Pd1/Familial Discord. Family that was critical, unsupportive, and interfered with independence.

Pd2/Authority Conflict. Rebellion against societal rules, beliefs of right/wrong that disregard societal norms, legal/academic difficulties.

Pd3/Social Imperturbability. Opinionated, socially confident, outspoken.

Pd4/Social Alienation. Isolated from others, feels poorly understood.

Pd5/Self-Alienation. Unhappy with self, guilt and regret regarding past behavior.

Scale 6. Paranoia

Pa1/Persecutory Ideas. Perceives world as dangerous, feels poorly understood, distrustful.

Pa2/Poignancy. Feels lonely, tense, hypersensitive, possibly high sensation seeking.

Pa3/Naiveté. Overly optimistic, extremely high moral standards, denial of hostility.

Scale 8. Schizophrenia

Sc1/Social Alienation. Feels unloved, mistreated, and possibly persecuted.

Sc2/Emotional Alienation. Depression, fear, possible suicidal wishes.

Sc3/Lack of Ego Mastery, Cognitive. Strange thoughts, sense of unreality, poor concentration and memory, loss of mental control.

Sc4/Lack of Ego Mastery, Conative. Depressed, worried, fantasy withdrawal, life is too difficult, possible suicidal wishes.

Sc5/Lack of Ego Mastery, Defective Inhibition. Sense of losing control of impulses and feelings, labile, hyperactive, cannot control or recall certain behaviors.

Sc6/Bizarre Sensory Experiences. Hallucinations, peculiar sensory and motor experiences, strange thoughts, delusions.

Scale 9. Hypomania

Ma1/Amorality. Selfish, poor conscience, manipulative; justifies amoral behavior by believing others are selfish and opportunistic.

Ma2/*Psychomotor Acceleration*. Restless, hyperactive, accelerated thoughts and behaviors, seeks excitement to reduce boredom.

Ma3/*Imperturbability*. Unaffected by concerns and opinions of others, denies feeling socially anxious.

Ma4/*Ego Inflation*. Unrealistic perception of abilities, resentful of demands placed on himself or herself.

Scale 0. Social Introversion

Si1/*Shyness*. Easily embarrassed, reluctant to initiate relationships, socially uncomfortable, shy.

Si2/*Social Avoidance*. Dislike and avoidance of group activities, parties, social activities.

Si3/*Self/Other Alienation*. Poor self-esteem, self-critical, low self-confidence, sense of ineffectiveness.

MMPI-2 CRITICAL ITEMS

An alternative to content analysis, other than scoring and interpreting actual scales, is to interpret the meanings of single items or clusters of items that seem, based on their content, to relate to different areas of psychopathology (depressed suicidal ideation, mental confusion, etc.); some items could represent serious pathology, regardless of how the person responded on the remainder of the inventory. These items have been referred to as *pathognomonic items*, *stop items*, or, more frequently, *critical items*. It has been assumed that the direction in which a person responds represents a sample of the person's behavior and acts like a short scale that indicates his or her general level of functioning. The critical items are most useful when clinicians look at the individual item content in relation to the specific types of information that the item reveals. This information might be used to guide further interviewing. However, some caution should be taken in interpreting these critical items, as they are both subject to an acquiescing response set (most items are keyed in the "True" direction) and to faking bad. They should not be considered to be scales but rather direct communications to the clinician about areas specific to the item content. A listing of critical items can be found in the MMPI-2 manual (Butcher et al., 2001); these items are typically scored by most computer-assisted programs.

Although lists of critical items have been included in the MMPI-2 manual (Butcher et al., 2001), clinicians should use these lists with caution in reference to adolescents. First, both normal adolescents and clinical populations of adolescents endorse, on average, twice the number of critical items as normal adults (Archer & Jacobson, 1993). In addition, normal adolescents and clinical populations endorse item frequencies about equally, thereby suggesting that the items themselves should not be used to differentiate between these two groups. This means that empirical attempts to develop critical item lists for adolescents might be quite difficult. As for the MMPI-2, clinicians should not treat the different clusters of critical items as rough scales to be interpreted. Rather, the individual item content should be used to develop specific

interview questions, and the relative deviancy of these items should be handled with appropriate tolerance.

MMPI-2 AND MMPI-A SUPPLEMENTARY SCALES

Since the initial publication of the MMPI, more than 450 new scales have been developed. Some of these have been developed for normals and are unrelated to pathology, such as dominance (*Do*) and social status (*St*). Other scales relate more directly to pathological dimensions, and often use the data from Hathaway and McKinley's original standardization sample or the more recent restandardization group. Scoring is possible only if the entire 567 MMPI-2 or 478 MMPI-A items are given. Although exact cutoffs for determining high scores have not been specified, they are generally $T = 65$. Scoring templates and profile sheets are available through Pearson Assessments. The scales that were selected for inclusion on this profile sheet are considered most useful, have been most extensively researched, or show promise in terms of future usefulness and/or are likely to be researched more extensively in the future. The next lists provide the names and interpretations surrounding scale elevations.

MMPI-2 Supplementary Scales

Al/Anxiety. High scores indicate that the person is upset, shy, retiring, insecure; has low self-confidence; is inhibited, uncertain, hesitant, conforming, under stress; and has extreme difficulty making decisions. Low scores indicate that the individual is extraverted, secure, relaxed, energetic, competitive, and generally has an absence of emotional difficulties.

R/Repression. High scorers tend to be submissive, overcontrolled, slow, clear thinking, conventional, formal, cautious; use denial and rationalization; and go to great lengths to avoid unpleasant interpersonal situations. Low scorers are likely to be dominant, enthusiastic, excitable, impulsive, self-indulgent, outspoken, and achievement oriented.

Es/Ego Strength. This scale assesses the degree to which a client is likely to benefit from psychotherapy, but it is probably specific to predicting the response of neurotic patients to insight-oriented therapy; it is probably not useful for other types of patients or other kinds of treatments. High scores suggest these persons can benefit from psychotherapy because they are likely to be adaptable and possess personal resources; have good reality contact; are tolerant, balanced, alert; have a secure sense of reality; will seek help in difficult situations; possess strongly developed interests; are persistent; can deal effectively with others; have a sense of personal adequacy; can easily gain social acceptance; and have good physical health. Low scores reflect general maladjustment. These people are likely to have low self-esteem and a poor self-concept; lack personal resources; feel insecure; are rigid and moralistic; have chronic physical problems; possess fears and phobias; are confused and helpless, have chronic fatigue; may be withdrawn and seclusive, inhibited; have personality rather than situational problems and poor work histories; and will therefore have difficulty benefiting from psychotherapy.

Dol/Dominance. Elevations indicate that the individual is self-confident, realistic, task-oriented; feels a sense of duty toward others; is competent to solve problems; is socially dominant, poised, and self-assured in working with groups; takes the initiative in relationships; possesses strong opinions; perseveres at tasks; and has a good ability to concentrate. The scale is useful and frequently used in personnel selection (e.g., police officer selection).

Rel/Responsibility. High scores suggest that the individual possesses high standards, a strong sense of justice and fairness, strong (even rigid) adherence to values; is self-confident, dependable, trustworthy. The scale is a general index of positive personality characteristics and is often useful in personnel screening.

Mt/College Maladjustment. High scores indicate general maladjustment among college students; they are likely to be worried, anxious, and procrastinate; they are pessimistic, ineffectual, somatize stress, and feel that, much of the time, life is a strain.

PK/Posttraumatic Stress Disorder Scale. High scores indicate emotional distress, depression, anxiety, sleep disturbances, guilt, loss of control over thinking, and a feeling of being misunderstood and mistreated by others. The scale does not determine that trauma has actually occurred but indicates that the symptoms reported are consistent with persons exposed to traumatic events; the existence of a trauma still needs to be determined through other means.

MDS/Marital Distress Scale. High scores indicate the person is experiencing marital distress; this scale is related more specifically to marital difficulties than either the FAM content scale or Scale 4 (both of which assess relationship difficulties not necessarily specific to marriage). *MDS* should be interpreted only for persons who are married, partnered, in a significant relationship, separated, or divorced.

Ho/Hostility Scale. High scores are characterized by being cynical, mistrusting, suspicious, unfriendly, and angry. However, they may not express their hostility in overt ways. They are likely to perceive others as being hostile and to blame others for their problems. As a result, they typically have low levels of social support. They have poor self-concepts and may be depressed, anxious, and experience somatic difficulties. They may have serious health problems.

O-H/Overcontrolled Hostility Scale. High scores suggest that the person is emotionally constricted, bottles up anger, and may overreact, possibly becoming physically or verbally aggressive; the aggressiveness usually occurs as rare incidents in a person who is otherwise extremely well controlled. The scale is most useful in understanding past behavior rather than predicting the likelihood of future hostility. Some persons who score high are not actively struggling to control dangerous hostility but are very well controlled and highly socialized. Thus, the scale is more directly a measure of persons who deny aggressive actions and are somewhat constricted; therapy, at least initially, might seem superficial and lacking in affect.

MAC-R/MacAndrew Alcoholism Scale-Revised. The MAC-R scale is best considered a measure of the potential for substance abuse. It differentiates between psychiatric outpatients with alcoholism and psychiatric outpatients who do not have alcoholism and identifies persons who are at risk of later developing alcohol-related problems. The potential to become involved in alcohol use is assessed rather than current alcohol use. In addition, the scale has difficulty differentiating alcohol abusers from other substance abusers. High scores on the MAC-R scale primarily suggest

actual or potential substance abuse but may also suggest extraversion, affiliation, confidence, assertiveness, risk-taking, sensation-seeking, past school behavior problems, the possibility of having experienced blackouts, and possible difficulties with concentration. Low scores are not only evidence against substance abuse but also may suggest introversion, conformity, and low self-confidence. Low scores in a person known to abuse substances suggest that the abuse is based more on psychological disturbance than typical addictive processes. The recommended raw score cutoff to indicate the initial point of drug and/or alcohol problems for males is 26 to 28; for females, it is a lower 23 to 25. The MAC-R is not particularly effective with African Americans and other racial and ethnic minority respondents. High scorers are likely to be extraverted, impulsive risk takers who will benefit from a group-oriented, confrontational treatment approach. Low scorers are more likely to be introverted, withdrawn, depressed risk avoiders who will be more likely to benefit from a supportive and relatively nonconfrontational treatment approach.

AAS/Addiction Acknowledgment Scale. High scores suggest a conscious awareness of and willingness to share information related to drug and/or alcohol-related problems. It is the most sensitive MMPI-2 scale for detecting substance abuse (Rouse, Butcher, & Miller, 1999; L. Stein, Graham, Ben-Porath, & McNulty, 1999). Low scores merely clarify that the person has not acknowledged these problems (although there is still the possibility that the person does have drug and/or alcohol-related difficulties).

APS/Addiction Potential Scale. High scores indicate that the person has a considerable number of lifestyle and personality factors consistent with those who abuse alcohol and/or drugs. The scale does not necessarily measure the extent of current use but more the potential for developing such problems. This means that if the APS (or MAC-R) is used to identify persons who are actually abusing substances, it is likely to result in a high number of false positives (Rouse et al., 1999). If the person scores in the normal to low range but history reveals a drug and/or alcohol problem, this problem is probably based primarily on psychological maladjustment (drug/alcohol use as self-medication) rather than a typical addictive pattern (harmful habits, peer group issues, physiological impact of the drug). This scale is quite similar to the MAC-R scale, but it uses more of the newer MMPI-2 item pool than the MAC-R. There is some indication that it measures the same factors as the MAC-R and may do so either as effectively (Rouse et al., 1999; L. Stein et al., 1999) or more effectively (Greene, Weed, Butcher, Arredondo, & Davis, 1992; Weed, Butcher, McKenna, & Ben-Porath, 1992).

GM/Masculine Gender Role. Persons who score high (both males and females) are likely to be self-confident, deny feeling afraid or worried, and be persistent in pursuing their goals; females scoring high are likely to be honest, unworried, and have a willingness to explore new things; high scores on GM with correspondingly low scores on GF indicate stereotypical male interests and orientations; high scores on both GM and GF suggest androgyny (the person has both masculine and feminine characteristics); low scores on GM along with high scores on GF suggest stereotypical feminine interests and orientation; low scores on both scales suggest an undifferentiated masculine/feminine orientation. This scale is still experimental and in need of further research.

GF/Feminine Gender Role. High scores suggest the endorsement of stereotypically feminine interests and orientations, and may also suggest religiosity and possibly abuse

of alcohol and/or nonprescription drugs; males scoring high may be hypercritical, express religiosity, avoid swearing but act bossy, and have a difficult time controlling their temper. This scale is still experimental and in need of further research.

MMPI-A Supplementary Scales

MAC-R/MacAndrew Alcoholism Scale. High scores suggest that the person is similar to others who have alcohol or drug problems; dominant, assertive, egocentric, self-indulgent, impulsive, unconventional; risk taker and sensation seeker; increased possibility of conduct disorder and legal difficulties. Low scores suggest that the person is dependent, conservative, avoids sensation-seeking activities, and is overcontrolled and indecisive.

ACK/Alcohol Drug Acknowledgment Scale. Adolescents who score high have a conscious awareness of and willingness to admit to alcohol- and/or drug-related problems; includes problem use, reliance on alcohol to cope or as a means of freely expressing feelings, and harmful substance abuse habits; friends or acquaintances may tell them that they have alcohol and/or drug problems; they may get into fights while drinking.

PRO/Alcohol Drug Proneness Scale. A high score suggests that the adolescent is prone to developing drug- and/or alcohol-related problems and school and home behavior problems. No obvious items related to drugs and alcohol are included on the scale; therefore, the scale measures personality and lifestyle patterns more consistent with alcohol- and drug-related problems. The scale does not so much measure current alcohol or drug use patterns (although they may still be present; quite similar to the MMPI-2 APS scale).

IMM/Immaturity Scale. High scorers are untrustworthy, undependable, boisterous; quickly become angry, are easily frustrated, may tease or bully others; are resistant, defiant, and likely to have a background of school and interpersonal difficulties.

A/Anxiety. General maladjustment, anxiety, distress, emotionally upset, experiences discomfort.

R/Repression. Submissive, conventional, works hard to avoid unpleasant or disagreeable situations.

MMPI-2-RF INTERPRETATION PROCEDURE

The MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011) represents a major revision of the MMPI-2. Indeed, it can be seen as both an alternative to as well as a stand-alone instrument. It is based on a subset of the MMPI-2 pool and utilizes the MMPI-2 normative sample. The validity scales have also been retained. In addition, many of the scales represent a refinement of the standard clinical scales. It has also been reduced in length to 388 items and, as such, represents a good alternative to the MMPI-2 when brevity is critical. At the core of the MMPI-2-RF are the restructured clinical scales and the Psychopathology Five (PSY-5) scales.

The next five steps are recommended for interpreting MMPI-2-RF profiles. As with the MMPI-2 and MMPI-A, these steps should be followed with awareness of the impli-

cations of age, culture, intellectual level, education, and level of functioning as well as the reason, motivation, and context of assessment. The discussion of the various scales and codes represents an integration and summary of both primary sources and the following MMPI-2-RF resources: Ben-Porath (2012); Friedman, Bolinsky, Levak, and Nichols (2014); Greene (2010); Handel and Archer (2008); Sellbom et al. (2006); and Tellegen et al. (2003).

Step 1. Score and Plot the Profile

Complete the scoring and plot the profile. While this can be accomplished by hand, it is recommended that evaluators use computer-assisted scoring, as hand scoring is more prone to error (Allard & Faust, 2000; R. Simons, Goddard, & Patton, 2002). Although there are 51 scales on the MMPI-2-RF, it is strongly recommended that every single one be scored.

Step 2. Determine Profile Validity

Assess the validity of the profile by noting the pattern of the validity scales. There are a number of indicators suggesting invalid profiles, which are clustered into three main areas: content nonresponsiveness, overreporting, and underreporting. Content nonresponsiveness includes the scales related to skipping items (CNS) and deliberately not attending to item content, including random responding (VRIN-r and TRIN-r). Overreporting includes scales related to the exaggeration of problems or psychopathology (F-r, Fp-r, Fs, FBS-r, and RBS). Underreporting includes scales related to a defensive test-taking style or the deliberate presentation of oneself as free of minor faults and overly virtuous (L-r and K-r). In addition to looking at the pattern presented by these validity scales, clinicians should consider the context of the assessment to determine whether a defensive, malingering, or inconsistent response style supports what is known about the client and his or her current situation. In particular, the examiner should determine whether over- or underreporting psychopathology or problems in general may lead to some type of gain (material, situational, or otherwise) for the client.

Step 3. Describe Symptoms, Behaviors, and Personality Characteristics from Substantive Scales

This step represents the core process in interpretation. Most of the substantive scales are considered elevated if $T > 64$ and highly elevated if $T > 79$, though for some scales a lower threshold is interpretable for highly elevated. Recommended ranges for interpretation are presented in the discussion of the substantive scales that follows. The substantive scales are configured into higher-order scales; restructured clinical scales; specific problem scales, which are further subcategorized into somatic/cognitive scales, internalizing scales, externalizing scales, and interpersonal scales; interest scales; and the personality psychopathology five (PSY-5) scales. Each of these is described in detail below. The higher-order and restructured clinical scales are presented in Table 7.1, and the specific problem scales, interest scales, and PSY-5 scales are presented in Table 7.3.

Table 7.3 MMPI-2-RF Specific Problem, Interest, and PSY-5 Scales

Name	Abbreviation	No. of items
Specific problem scales		
Malaise	MLS	8
Gastrointestinal complaints	GIC	5
Head pain complaints	HPC	6
Neurological complaints	NUC	10
Cognitive complaints	COG	10
Suicidal/death ideation	SUI	5
Helplessness/hopelessness	HLP	5
Self-doubt	SFD	4
Inefficacy	NFC	9
Stress/worry	STW	7
Anxiety	AXY	5
Anger proneness	ANP	7
Behavior-restricting fears	BRF	9
Multiple specific fears	MSF	9
Juvenile conduct problems	JCP	6
Substance abuse	SUB	7
Aggression	AGG	9
Activation	ACT	8
Family problems	FML	10
Interpersonal passivity	IPP	10
Social avoidance	SAV	10
Shyness	SHY	7
Disaffiliativeness	DSF	6
Interest scales		
Aesthetic-literary interests	AES	7
Mechanical-physical interests	MEC	9
PSY-5 scales		
Aggressiveness–revised	AGGR-r	18
Psychoticism–revised	PSYC-r	26
Disconstraint–revised	DISC-r	20
Negative emotionality/neuroticism–revised	NEGE-r	20
Introversion/low positive emotionality–revised	INTR-r	20

During the interpretive process, it is important that in addition to noting the meanings of individual scales, the clinician should look at the overall pattern of scales and how they work together for the individual being evaluated. That is, while a high *T* score on the Low Positive Emotions scale (RC2) may indicate signs of depression, understanding this in relation to the Introversion/Low Positive Emotionality—Revised scale (INTR-r), which relates to general personality characteristics like pessimism, insecurity, and self-criticism can help elucidate the meaning of the high RC2 score. While currently feeling sad, disengaged socially, and introverted, this may be characteristic of a typical long-standing proneness toward depression (with high INTR-r), or it may be situationally and context-dependent, going against the client's typically outgoing and positive outlook on life (with low INTR-r). While 2-point (and other) code types are not typically used on the MMPI-2-RF, some clusters of scales relate to each other in terms of general content.

Substantive Scale Clusters

When interpreting the substantive scales, clusters of rationally and conceptually complementary scales can help organize the data to better characterize the specific individual being evaluated. Conceptually clustering scales is important so as not to provide a generic description of a person based on a single scale, which could easily characterize many different individuals. Understanding that different aspects of personality, emotional, and behavioral functioning vary by person (and indeed by situation and context) allows the evaluator the ability to consider multiple scales together to add nuance to the description of the individual. The next cluster areas and general interpretive strategies are basic approaches to help guide hypothesis generation around specific areas. While these guidelines will serve to alert clinicians to specific areas, clinicians will still need to investigate these areas in far more depth by consulting relevant scale descriptors and patterns between scales.

Somatic/Cognitive Dysfunction. Multiple scales provide information on somatic and cognitive problems. Included are RC1 (Somatic Complaints), MLS (Malaise), GIC (Gastrointestinal Complaints), HPC (Head Pain Complaints), NUC (Neurological Complaints), and COG (Cognitive Complaints). The RC1 scale represents overall pre-occupation with health concerns and the likelihood of having physical complaints in reaction to stress, and MLS relates to overall how debilitated the individual is presenting him- or herself. In contrast, the GIC, HPC, and NUC scales focus on the specific types of physical problems the person is presenting. The COG scale often represents somewhat of an “outcome” of other problems, focusing on difficulties with attention, memory, and concentration, among other cognitive complaints.

Emotional Dysfunction. Not surprisingly, given the general scope and purpose of the MMPI-2-RF, many scales relate to reports of emotional issues. Included are EID (Emotional/Internalizing Dysfunction), RCd (Demoralization), RC2 (Low Positive Emotions), RC7 (Dysfunctional Negative Emotions), SUI (Suicidal/Death Ideation), HLP (Helplessness/Hopelessness), SFD (Self-Doubt), NFC (Inefficacy), STW (Stress/Worry), AXY (Anxiety), ANP (Anger Proneness), BRF (Behavior-Restricting Fears), MSF (Multiple Specific Fears), NEGE-r (Negative Emotionality/Neuroticism—revised), and INTR-r (Introversion/Low Positive Emotionality—revised). The acute emotional problems presented by the majority

of these scales (see the specific interpretation of each scale below) can be interpreted within the context of the two relevant PSY-5 scales, the NEGE-r and INTR-r, which represent more personality tendencies toward emotional functioning. For example, elevations on NEGE-r may represent a proneness toward catastrophizing, which may underlie (perhaps in combination with current circumstances) elevations on STW and AXY. The NEGE-r and INTR-r scales can help clinicians determine whether any current emotional problems are more likely reactive to some outside stressor or circumstance (with low NEGE-r and low INTR-r), and thus likely more transient in nature, or if they are more likely chronic and dispositional in nature (with high NEGE-r and/or INTR-r), which would be more difficult to treat and overcome.

Thought Dysfunction. In addition to the higher-order scale of the same name, several scales can help determine the likely presence of disturbance in thinking. Included are THD (Thought Dysfunction), RC6 (Ideas of Persecution), RC8 (Aberrant Experiences), and PSYC-r (Psychoticism–revised). Each of these scales adds evidence of problems with reality testing, unrealistic thinking, unusual perceptual experiences, and, in the case of RC6, specific paranoid and persecutory ideas. Although this cluster of scales is clearly related to psychotic and disorganized thought processes and cluster A personality disorders, some neurological disorders will also elevate some of these scales, which include items related to bizarre perceptual abnormalities and confused thinking. Elevation on RC6 would be evidence of psychotic processes, whereas elevation on RC8 could indicate psychosis or a neurological problem.

Behavioral Dysfunction. Behavioral problems can take many forms, and this cluster of scales can help determine how intrusive they likely are, as well as the specific types of problems. Included are BHX (Behavioral/Externalizing Dysfunction), RC4 (Antisocial Behavior), RC9 (Hypomanic Activation), JCP (Juvenile Conduct Problems), SUB (Substance Abuse), AGG (Aggression), ACT (Activation), AGGR-r (Aggressiveness–revised), and DISC-r (Disconstraint–revised). Broadly, this cluster of scales evaluates acting out behaviors and some underlying mechanisms related to them. Although BHX may characterize the extent of the tendency to act out, the specific type of acting out can be reflected in the other scores, such as whether it is related to using alcohol or other drugs (SUB, which in and of itself can be an acting out behavior), interpersonal aggression (AGG), rule-breaking behaviors (RC4), or others. Some of the scales can help the clinician understand why the individual may be acting out, such as a heightened level of excitability (ACT) or sensation-seeking (RC9), general impulsivity (DISC-r), use of alcohol or other drugs (SUB, which can disinhibit individuals and lead to acting-out behavior), a general demeanor of social dominance (AGGR-r), or other mechanisms.

Interpersonal Functioning. The cluster of interpersonal scales, in addition to some of the Behavioral Dysfunction scales discussed earlier, can help the clinician understand the ways in which the client interacts with others, in addition to how he or she may interact with a therapist in treatment. Included in this cluster are FML (Family Problems), RC3 (Cynicism), IPP (Interpersonal Passivity), SAV (Social Avoidance), SHY (Shyness), DSF (Disaffiliativeness), and INTR-r (Introversion/Low Positive Emotions–revised). Although some of the scales look more at interpersonal behaviors (IPP and SAV), others look more at interpersonal attitudes and feelings (SHY, DSF, and INTR-r). The FML scale provides extra information about possible reasons

for any interpersonal dysfunction (if FML represents responses related mostly to family of origin), as well as the current context in which much of the interpersonal dysfunction may play out (if FML represents responses related to current family discord; see section titled “*FML/Family Problems*” later in this chapter).

Interests. Although not comprehensive, two scales represent the general types of activities in which the individual is interested. These two scales were derived from the original Scale 5/Mf (Masculinity/Femininity). Included in this cluster are AES (Aesthetic-Literary Interests) and MEC (Mechanical-Physical Interests). Fairly straightforward to interpret (see sections on each of these scales later in this chapter), individuals who score low on both scales are disengaged from the world around them, generally lacking interest in activities. Clinicians should check other scales related to emotional dysfunction to see if this lack of interest is related to, among other possibilities, a current depression or characteristically low activation.

Step 4. Provide Diagnostic Impressions

Like the MMPI-2, although the MMPI-2-RF does not and should not provide direct diagnoses, it contributes considerable information relevant to diagnostic formulations. Potential *DSM-5* diagnosis considerations for elevations on individual scales are included below. Clinicians should consider these, along with additional available information, to help make an accurate diagnosis. In some contexts and for some types of referral questions, formal diagnosis will be relevant; but for other contexts and referral questions, formal diagnosis will be neither required nor appropriate (e.g., employment screening).

Step 5. Elaborate on Treatment Implications and Recommendations

Often, one of the most valuable services a practitioner can provide is to predict the client’s likelihood of benefiting from interventions. Doing this typically means elaborating on the person’s strengths and weaknesses, level of defensiveness, ability to form a treatment relationship, predicted response to psychotherapy (including different types of psychotherapy), antisocial tendencies, and level of insight. Much of this information is summarized at the ends of the subsections on scale elevations. If clinicians do extensive work with specific types of clients, they might need to expand on their knowledge relating to types and outcome of treatments by referring to the extensive research base that is available (e.g., chronic pain, substance abuse). Treatment responsiveness might be further extended into providing suggestions for tailoring specific interventions for client profiles and types of problems. While not much has been reported on outcomes of tailored interventions for specific scale elevations yet with the MMPI-2-RF, a useful resource in the treatment planning process is Maruish’s (2004) *Use of Psychological Testing for Treatment Planning and Outcome Assessment*.

MMPI-2-RF VALIDITY SCALES

The MMPI-2-RF developed its group of validity scales using two major, overarching principles. First, theoretically, three major threats to the validity of the test were

identified: content nonresponsiveness (the client not attending to the content of the items of the test), overreporting, and underreporting. The development of the validity scales for the MMPI-2-RF would need to ensure that each of these three threats was comprehensively evaluated. The second principle that guided the development of the new validity scales was that the validity scales from the MMPI-2 are both conceptually and empirically strong, though with some psychometric problems. Thus, the new scales are based largely on the scales from the MMPI-2, but with specific important tweaks to most. Tellegen and Ben-Porath (2008/2011) provide details on the development of each of these revised validity scales in the MMPI-2-RF manual.

The Cannot Say (CNS) “scale,” based on the original ?/Cs “scale,” remains merely a count/percentage of items that went unanswered (or answered both true and false, although this is rare) by the respondent. It functions in the same way as the original scale (see detailed description later in this chapter). The revised Variable Response Inconsistency and True Response Inconsistency (VRIN-r and TRIN-r, respectively) scales are also based on, and function like, the corresponding VRIN and TRIN scales on the MMPI-2. However, these scales were revised so that there were no item overlaps on the two scales (which is a concern on the MMPI-2, as elevation in one scale can necessarily elevate the other, because of item overlap). The overreporting indicators include revised versions of Infrequent Responses (F-r), Infrequent Psychopathology Responses (Fp-r) and Symptom Validity (FBS-r), as well as the added scales Infrequent Somatic Responses (Fs) and Response Bias Scale (RBS). The guiding principles for the revision of the three scales from the MMPI-2 included ensuring that the criterion for low endorsement of items was updated to the current normative sample (instead of being based on the original normative sample of the MMPI), as well as reducing (nearly eliminating) item overlap between the scales. The two additional overreporting scales, the Fs and the RBS, were developed and added after revision of the other scales, and they consequently have a few items overlapping with the other F scales (though not many). The Fs scale was developed by Wygant, Ben-Porath, and Arbisi (2004) to identify uncommonly endorsed somatic symptoms, even in large medical samples. The RBS was developed by Gervais, Ben-Porath, Wygant, and Green (2007) by identifying items that significantly predicted which disability claimants passed or failed other widely used and validated tests of malingering.

Because identifying underreporting of problems is a more difficult task than identifying overreporting, and because of mixed findings in the literature about the L, K, and S scales on the MMPI-2, the development of the underreporting validity scales took a different approach. Baer and Miller (2002) evaluated the utility of two scales of underreporting from the original MMPI, compared to the three MMPI-2 scales, and found them to be useful, even outperforming the MMPI-2 scales. These scales were the Positive Malingering Scale (Mp; Cofer, Chance, & Judson, 1949) and the Wiggins Social Desirability Scale (Wsd; Wiggins, 1959). When items of all five scales were examined with factor analysis (in personnel selection, simulated underreporting, and clinical samples), two factors emerged (Tellegen & Ben-Porath, 2008/2011; see also Bagby & Marshall, 2004). The two new/revised underreporting scales were constructed by selecting items that substantially loaded onto each of these two factors, without loading significantly onto the other, and emerged as analogous to the L and K scales on the MMPI-2. The Uncommon Virtues (L-r) and Adjustment Validity (K-r) scales on the

MMPI-2-RF are both associated with significant underreporting of symptoms, problems, and faults.

Content Nonresponsiveness Scales

CNS/Cannot Say. As in the MMPI-2, the CNS scale is not actually a formal scale but merely represents the number of items that are unscorable, either left unanswered or answered both True and False on the profile sheet. The usefulness of noting the total number of unscorable questions is to provide one of several indices of a protocol's validity. If 15 or more items are unscorable, the protocol's validity should be seriously questioned. This is simply because an insufficient number of items have been responded to, which means less information is available for scoring the scales. Thus, less confidence can be placed in the results. To minimize the number of CNS responses, the client should be encouraged to answer all questions, and items with both responses should be queried to determine which response should be scored.

High Number of CNS (15+)

- Difficulties with reading, psychomotor retardation, indecision, confusion, or extreme defensiveness (consistent with severe depression, obsessional states, extreme intellectualization, or unusual interpretations of the items).
- Legalistic overcautiousness or a paranoid condition.
- Perception that the unanswered items are irrelevant.

VRIN-r/Variable Response Inconsistency. The VRIN-r scale includes 53 pairs of selected items that would be expected to be answered in the same direction if the person is approaching the testing in a valid way, attending appropriately to the item contents. Unlike the MMPI-2, each pair of items is directionally similar in content; respondents would be expected to respond either True to each item in a pair or False to each item in a pair. If a person answers in the opposite direction, then it indicates an inconsistent response and is therefore scored as 1 raw score on the VRIN-r scale. Inconsistent responding may reflect motivated uncooperativeness (randomly responding without reading the items) or unintentional problems, such as difficulty with language or understanding the items.

Moderate VRIN-r ($T = 70-79$)

- Somewhat inconsistent responding; profile should be interpreted with caution, and other validity scales should be evaluated for further evidence of problematic responding.
- May reflect carelessness, reading or comprehension difficulties.

High VRIN-r ($T > 79$)

- Indiscriminate responding; profile should be considered invalid and should not be interpreted.

TRIN-r/True Response Inconsistency. The TRIN-r scale also includes pairs of items; however, only pairs with opposite contents are included. This means there would be two ways for a person to obtain a response that would be scored on the TRIN-r scale. A “True” response to both items would indicate inconsistency and would therefore be scored as plus 1 raw score point. A “False” response to both pairs would also indicate inconsistency but would be scored as minus 1 point. Final VRIN-r *T* score is transformed so that it is always in the positive direction, such that higher scores are related to more problematic responding (regardless of whether the “points” were accumulated based on True-True pairs or False-False pairs). The score is reported with the letters “T” and “F” to distinguish those respondents who primarily responded True to all items or False to all items, respectively. That is, an individual with a *T* score of 80 on the TRIN-r could have a score of 80T, indicating that he or she tended to answer True to most items, or 80F, indicating that he or she consistently answered False.

Very High ($T > 79$) or Moderate (70–79)

- Person is indiscriminately answering “True” to the items (acquiescence or yea-saying) or indiscriminately answering “False” to the items (nonacquiescence or naysaying).
- Uncooperative approach to test-taking. Very high scores render the test uninterpretable.

Overreporting Scales

F-r/Infrequent Responses. The F-r scale measures the extent to which a person answers in an atypical and deviant manner. The scale items were selected based on their endorsement by less than 10% of the current normative sample. Thus, from a statistical definition, they reflect nonconventional responding. The items do not cohere around any particular trait or syndrome. High scores indicate the examinee is answering in a scorable direction to a wide variety of unusual characteristics. High scores can be used as a general indicator of pathology. In particular, high scores can reflect unusual feelings caused by some specific life circumstance to which the person is reacting. This might include grieving, job loss, or divorce. However, extremely high scores are infrequent even in populations with severe psychopathology; thus, the individual is likely “faking bad,” which will serve to invalidate the protocol. No exact cutoff score is available to determine whether a profile is invalid or is accurately reflecting pathology. It is important to note that elevations on F-r should first be evaluated in relation to the VRIN-r and TRIN-r scales. That is, random responding or all-true or all-false responding may elevate this scale, not because of pathology or faking bad but because of lack of attending to item content. Only when it is determined that these reasons are not the cause of elevations on F-r should the F-r be interpreted.

High Scores on F-r (approximately $T > 99$; very high $T > 119$)

- Potentially invalid profile; should be considered definitely invalid at very high scores.

- Possibly caused by clerical errors in scoring, false claims by the client regarding symptoms, resistance to testing, malingering.

Moderate Scores ($T = 79-99$)

- Attempt to draw attention to distress as a cry for help (and are in need of assistance).
- May reflect true psychopathology and distress or exaggeration and overreporting of symptoms.

Fp-r/Infrequent Psychopathology Responses. Because the F-r scale is typically elevated among psychiatric patients, it is often difficult to differentiate between persons with true psychopathology and those who have some psychopathology but are nonetheless faking bad (or are just purely faking bad). This is particularly true if the psychopathology is quite severe. The history of the person (e.g., degree of preexisting psychopathology) and context of the referral (e.g., possible gain for faking bad) can often be quite useful in making this distinction. To further assist with this differentiation, the Fp-r scale is a set of 21 items that were infrequently endorsed even by psychiatric inpatients. (In contrast, the F-r scale was developed from infrequently endorsed questions by the normative sample.) As with the F-r, elevations on Fp-r should first be evaluated in relation to the VRIN-r and TRIN-r scales. That is, random responding or all-true or all-false responding may elevate this scale, due to the lack of attending to item content. Only when it is determined that these reasons are not the cause of elevations on Fp-r should the Fp-r be interpreted.

High Scores on Fp-r (approximately $T > 69$; very high $T > 99$)

- High probability of faking or exaggerating psychopathology, even among psychiatric patients.
- Very high scores should render the scores on the Substantive Scales uninterpretable.

Fs/Infrequent Somatic Responses. The Fs scale includes 16 items that relate to somatic complaints that were endorsed infrequently even by patients being treated for medical diseases and problems. Like the F-r and Fp-r scales, mild elevations can reflect genuine problems (in this case, physical complaints), but higher elevations likely reflect exaggeration or overreporting of somatic symptoms. There are many reasons that individuals may be motivated to overreport somatic and physical symptoms, including secondary gain (as in a disability or personal injury claim), medication seeking, and Factitious Disorder. As with the F-r and Fp-r, elevation on the Fs should first be evaluated within the context of the VRIN-r and TRIN-r, because if either of these scales is elevated, the response strategy (highly variable or overwhelmingly True) may elevate Fs, rather than responses to the actual Fs item content.

High Scores on Fs (approximately $T > 79$; very high $T > 99$)

- High probability of faking or exaggerating somatic symptoms, with endorsement higher than individuals with genuine medical problems.

- Data from outside of the MMPI-2-RF should be evaluated as context for interpreting this scale, such as genuine medical problems, potential secondary gain from disabilities, and so on.
- Somatic scales should be interpreted with caution (and considered invalid when *Fs* is very high). Included are Somatic Complaints (RC1), Malaise (MLS), Gastrointestinal Complaints (GIC), Head Pain Complaints (HPC), and Neurological Complaints (NUC).

FBS-r/Symptom Validity. Like the Fake Bad Scale (FBS) from the MMPI-2, the FBS-r is used to evaluate a cluster of noncredible symptoms, mostly somatic and cognitive, that reflect overreporting among those in the context of civil litigation. The FBS-r contains 30 items and, again, should be interpreted within the context of elevated VRIN-r and TRIN-r, just as the preceding scales are.

High Scores on FBS-r ($T > 79$; very high $T > 99$)

- High possibility of malingering, especially at very high scores.
- Data from outside of the MMPI-2-RF should be evaluated as context for interpreting this scale, such as genuine medical problems, potential secondary gain from disabilities, and so on.
- Somatic and cognitive scales should be interpreted with caution (and considered invalid when FBS-r is very high). Included are the same somatic scales as with *Fs* (see the earlier section titled “*Fs*/Infrequent Somatic Responses”), in addition to the Cognitive Complaints (COG) scale.

RBS/Response Bias Scale. The RBS scale contains 28 items used to detect a response bias toward overreporting problems, similar to the other overreporting scales. The primary difference of the RBS is that it was developed specifically within the context of disability and personal injury claimants and correlated highly with measures of cognitive and memory malingering. That is, it is effective at evaluating motivated overreporting among those with potential for gain. As such, it is a useful addition to the overreporting validity scales. As with the other overreporting scales, the RBS should be interpreted within the context of elevated VRIN-r and TRIN-r, as a random or otherwise nonattending response style may elevate the RBS.

High Scores on RBS ($T > 79$; very high $T > 100$)

- Possibility of overreporting problems. High scores may reflect individuals with significant emotional disturbance, rather than manipulative overreporting for gain. Very high scores are unusual even for those with severe emotional disturbance.
- Data from outside of the MMPI-2-RF should be evaluated as context for interpreting this scale, such as genuine medical problems, potential secondary gain from disabilities, etc.
- The Cognitive Complaints (COG) scale should be interpreted within the context of scores on the RBS.

Underreporting Scales

L-r/Uncommon Virtues. The L-r scale consists of 14 items that indicate the extent to which a client is attempting to describe him- or herself in an unrealistically positive manner. High scorers describe themselves in an overly idealized manner. The items consist of descriptions of relatively minor flaws to which most people are willing to admit. Thus, persons scoring high on the L-r scale might state that they never get angry or that they like everyone they meet. It is important to note that individuals raised with or currently holding strict, traditional values may have elevations on this scale, simply because they are responding the way they think they *should* respond. For those with elevations on L-r, the Substantive Scales may represent underestimates of actual problems. As with all the other validity scales, elevation on L-r should be interpreted within the context of any elevation on VRIN-r or TRIN-r, as nonattentive responding may elevate this scale without having anything to do with the actual content of the items.

High Scores on L-r ($T > 79$)

- Person is describing self in an overly favorable light due to conscious deception.
- Person is describing self in an overly favorable light due to an unrealistic view of him- or herself; may be inflexible, unoriginal, and unaware of the impressions he or she makes on others; perceives world in a rigid, self-centered manner.
- Extremely high scores might be due to conscious deception by antisocial personalities.
- Do not interpret absence of elevation on the Substantive Scales.
- Interpret elevations on the Substantive Scales as likely underestimates of problems.

Moderate Scores on L-r ($T = 65-79$)

- May be underreporting problems and issues.
- May be due to traditional upbringing and strict moral and behavioral values.
- Interpret Substantive Scales with caution, as they may be underreporting difficulties.

K-r/Adjustment Validity. The K-r scale includes 14 items (all of which were included on the original MMPI-2 K scale) that assess self-reported level of adjustment. Individuals can present themselves as extremely well adjusted for two primary reasons. First, they may in fact be extremely well adjusted, even better adjusted than most people. Information from outside the test and testing situation can be extremely useful in determining if elevations on K-r reflect this. Second, they may present themselves in an overly favorable manner, underreporting any problems or difficulties deliberately. Similar to the L-r scale, elevations on K-r indicate that the Substantive Scales may be underestimates of actual problems; those scales without elevation may be hiding pathology, and those with elevation may be underestimating it. Also like the L-r and the other validity scales, elevation in VRIN-r and TRIN-r can represent response styles that may elevate K-r in a manner unrelated to item content. K-r should only be interpreted if VRIN-r and TRIN-r are not significantly elevated.

High Scores on K-r ($T > 69$)

- Person is describing self as extremely well adjusted due to conscious deception.
- Person is describing self as extremely well adjusted due to an unrealistic view of him- or herself; may be inflexible, unoriginal, and unaware of the impressions he or she makes on others; perceives world in a rigid, self-centered manner.
- Do not interpret absence of elevation on the Substantive Scales.
- Interpret elevations on the Substantive Scales as likely underestimates of problems.

Moderate Scores on K-r ($T = 65-79$)

- May genuinely be extremely psychologically well adjusted.
- May be underreporting problems related to coping with stress or feeling distressed.
- Interpret Substantive Scales with caution, as they may be underreporting difficulties.

MMPI-2-RF HIGHER-ORDER SCALES

The three higher-order scales, which emerged from factor analyses (see Tellegen & Ben-Porath, 2008/2011), represent broad domains in personality and emotional functioning and describe broad, overall functioning in these three areas. These areas—emotions, thinking, and behavior—are broad-based, global areas of functioning and are susceptible to within-domain variation. That is, average scores on any one of these domains may represent good adjustment, or they may represent highly variable scores on the more detailed subdomains that make them up. For example, an individual may score average on the Emotional/Internalizing Dysfunction (EID) scale, and this may be due to good emotional adjustment. However, it could also represent extremely high elevation on one subdomain, such as Multiple Specific Fears (MSF), with low scores on the others. This may be the case for an individual who has a specific phobia but does not struggle with other anxieties, depression, self-doubt, or other emotional problems. Thus, elevation on these higher-order scales will indicate general problems in each of the three areas, but lack of elevation may not necessarily mean entirely positive adjustment. Interpretation should always be followed by evaluation of the other Substantive Scales that relate to the construct.

Emotional/Internalizing Dysfunction (EID)

The EID scale includes 41 items and evaluates general emotional functioning. Though lack of elevation is not clearly interpretable, in general, elevation on EID relates to greater emotional difficulty and inner turmoil. The EID scale often relates to an individual's subjectively felt distress (including anxious or depressed feelings) when responses to the MMPI-2-RF are open and honest.

High Scores on EID ($T > 64$; very high $T > 79$)

- Emotional distress, often related to anxiety and depression.

- Likely a clear awareness that individuals are in distress and not coping well with life stressors.
- At very high scores, they are likely to feel that they are currently in a significant crisis state.

Related Scale Elevations

- Specific details as to the nature of the emotional distress can be found in the other emotional scales.
- Included are RCd, RC2, RC7, SUI, HLP, SFD, NFC, STW, AXY, ANP, BRF, MSF, NEGE-r, and INTR-r.

Treatment Implications

- Often motivated for treatment, as their current distress leads to a discomfort and acknowledgement of the need for help.

Thought Dysfunction (THD)

The THD scale includes 26 items and evaluates multiple problems with thinking. While again lack of elevation is not clearly interpretable, in general, elevation on THD relates to greater dysfunction in thought process. Individuals with elevation on THD are more likely to have more pervasive disordered thinking that will interfere with their everyday functioning.

High Scores on THD ($T > 64$; very high $T > 79$)

- Thought dysfunction, which may be related to problematic perceptual experiences (e.g., illusions and hallucinations), belief systems and thoughts (e.g., delusions), and thought processes (e.g., confusion).
- Not necessarily associated with level of insight.
- At very high scores, the individual's disordered thinking is serious enough to severely interfere with his or her functioning.

Related Scale Elevations

- Specific details as to the nature of the thought dysfunction can be found in the other thinking scales.
- Included are RC6, RC8, and PSYC-r.

Treatment Implications

- Should be evaluated for the need for higher levels of treatment, including possible inpatient stabilization and psychopharmacological intervention.

Behavioral/Externalizing Dysfunction (BXD)

The BXD scale includes 23 items and evaluates multiple problems with behavioral disconstraint and acting out. While again lack of elevation is not clearly interpretable, in general, elevation on BXD relates to greater behavioral problems. Individuals with

elevation on BXD are likely to have gotten or to get into trouble for their acting out behavior.

High Scores on BXD ($T > 64$; very high $T > 79$)

- Problematic lack of behavioral control, which has likely led (and can lead) to difficulties, including interpersonal, legal, educational, and occupational problems.
- Behavioral acting out characterizes a primary mechanism for coping with stress and other difficulties.
- At very high scores, the individual's behavior problems are serious enough to severely interfere with his or her functioning.

Related Scale Elevations

- Specific details as to the nature of the tendency to act out can be found in the other behavior scales.
- Included are RC4, RC9, JCP, SUB, AGG, ACT, AGGR-r, and DISC-r.

Treatment Implications

- Unlikely to engage in traditional treatment, unless motivated by external factors.
- Once engaged, likely to act out, including being noncompliant with treatment or confrontational with the provider.

MMPI-2-RF RESTRUCTURED CLINICAL SCALES

As noted previously, the MMPI clinical scale items were selected based on their ability to distinguish clinical groups from normals. The scale items were not designed to differentiate various clinical groups from each other. The result has been that the MMPI-2 clinical scales are good at identifying that psychopathology is present but not what the psychopathology is. In order to counter this difficulty, many strategies have been developed to help clinicians make more nuanced distinctions among various scale elevations on the MMPI-2. These strategies have included the content scales, Harris and Lingoes subscales, supplementary scales, and critical items. In 2003, a set of restructured clinical scales was made available to help isolate the core features of the clinical scales (Tellegen et al., 2003). The first step was to use factor analysis to develop a general demoralization scale (RCd/Demoralization). By extracting all items in the demoralization scale from any of the clinical scales, initial (or “seed”) scales were developed that more closely assessed the core features of the clinical scales. These scales were then further refined to develop the final Restructured Clinical scales (see reviews and debate in a special series of the *Journal of Personality Assessment*, Meyer, 2007).

Demoralization (RCd)

The RCd scale evaluates broad emotional dissatisfaction and dysphoria, including discouragement, pessimism, poor self-esteem, insecurity, a sense of failure, emotional discomfort, poor coping ability, helplessness, interpersonal sensitivity, depression, anxiety, and presence of somatic symptoms. The scale relates to the general emotional

burden people feel in life, overwhelmed by stressors both from within and outside of themselves. The specific quality of the general dysphoric demoralization can be evaluated using the Specific Problems and Psy-5 scales.

High Scores on Scale RCd ($T > 64$; very high $T > 79$)

- Emotional turmoil most often related to dissatisfaction and unhappiness.
- Overwhelmed with life.
- Pessimistic and hopeless about the future, as well as feeling helpless, ineffective, and insecure.
- Anxious and depressed, including ruminative negative thoughts about the future and past.

Related Scale Elevations

- Evaluate SUI and HLP for risk for suicidal ideation and self-harm.
- Anxiety should be evaluated using the STW, AXY, BRF, MSF, and NEGE-r scales.
- Depression should be evaluated using the SFD, NFC, ANP, and INTR-r scales.
- The interpersonal scales (FML, IPP, SAV, SHY, and DSF) can provide a context for the demoralization, including possible causes and outcomes.

Treatment Implications

- Evaluate for potential suicidal risk.
- Emotional distress can motivate individuals to engage in treatment, and alleviation of these symptoms early in treatment can solidify the therapeutic commitment.

Somatic Complaints (RC1)

The RC1 scale evaluates the presence of significant health concerns, whether they are based in reality or not. While some elevation may represent some genuine medical difficulties, with higher scores, a significant psychological component is much more likely. The scale looks both at the somatization of physical complaints and the worry related to physical health, which may include weakness, fatigue, gastrointestinal upset, and chronic pain. Individuals with elevations on RC2 often deny psychological explanations for their physical worries or symptoms.

High Scores on Scale RC1 ($T > 64$; very high $T > 79$)

- Multiple somatic complaints, possibly due (in part or wholly) to an actual medical condition.
- Prone to the physical expression of emotional stress.
- Very high scores increase the likelihood that there is a somatizing component to the individual's presentation, which may include exaggeration and catastrophization of minor symptoms.
- Higher scorers are less likely to acknowledge that their physical problems may be psychologically based.

Related Scale Elevations

- Evaluate MLS, GIC, HPC, NUC, and COG for different specific types of somatic distress.
- Conversion Disorder should be examined, especially with low RC3 and SHY scales.

Treatment Implications

- May reject psychological interventions if they imply that there is a psychological component to their physical difficulties.
- Should consider using more bodily based interventions, especially at the beginning of treatment, such as relaxation techniques, deep breathing, progressive muscle relaxation, or hypnosis.

Low Positive Emotions (RC2)

The RC2 scale evaluates the extent to which respondents deny positive emotions and experiences. Although common in depression, low positive emotions are not directly related to affective disorders and can be related to other psychopathology. Traits that are associated with low positive emotions are being withdrawn and isolated, passive, self-critical, and having insufficient energy to deal with life challenges, which can make accomplishing tasks difficult. Individuals are often bored, with little ability to experience pleasure, and can be highly interpersonally sensitive, which can relate to difficulty becoming engaged with people. Because only responses of false to the items add to this scale, an individual who is responding false in a fixed way (not attending to item content, as measured by the TRIN-r) will have a falsely inflated score on RC2.

High Scores on Scale RC2 ($T > 64$; very high $T > 79$)

- Lacking in energy, vitality, and an engagement in life and relationships.
- Very little enjoyment in activities or relationships.
- Self-critical, with poor expectations of success and accomplishment.
- Possibly depressed.

Related Scale Elevations

- While no other scales are directly related to the content of the RC2 scale, HLP, SFD, and NFC should be evaluated for their relation to possible depression.
- The SUI scale should also be evaluated in terms of elevation on RC2.

Treatment Implications

- Evaluate for potential depressive disorder and suicidal risk.
- Unlike RCd, which relates to feelings of distress, the lethargy related to RC2 elevations may deter individuals from seeking out or engaging in treatment.
- Perhaps consider the use of antidepressant medication in conjunction with therapy.

Cynicism (RC3)

The RC3 scale evaluates a generally negative view of human nature, believing that others are uncaring and motivated by self-gain. An important distinction is the cynical belief about all people, as opposed to the feeling that they themselves are being negatively targeted (as would be measured by RC6). Because high scorers believe that self-interest is the primary motivation of others, they tend to be uncaring and untrustworthy, readily willing exploit others. In contrast, low scorers are likely to be gullible, naive, and trust others too easily.

High Scores on Scale RC3 ($T > 64$; very high $T > 79$)

- Cynical about others' intentions and mistrustful of others in general.
- Uncaring and hostile toward others.
- Typically have negative interpersonal interactions and distant (at best) relationships.

Related Scale Elevations

- Evaluate the interpersonal scales (FML, IPP, SAV, SHY, and DSF) for attitudes, values, and behaviors related to RC3. Many of these could explain these individuals' cynical nature (e.g., elevated FML may reflect a tumultuous and problematic early family experience), as well as some possible outcomes of it (e.g., elevated DSF may reflect a distaste for interaction and relationships stemming from a cynical attitude of others).
- Evaluate ANP and AGG to see if cynicism is related to anger and aggressiveness toward others.

Treatment Implications

- Likely to question the motivations of the therapist, which can hinder the development of a therapeutic alliance.

Antisocial Behavior (RC4)

The RC4 scale evaluates a range of traits related to antisocial personality disorder, including a history of (and current) breaking rules, lying, cheating, stealing, and abusing substances. Conflicts with family (past and present) also factor into this scale. Individuals with elevation on RC4 are likely to have had legal difficulties, poor achievement in general, and possibly a history of sexual acting out.

High Scores on Scale RC4 ($T > 64$; very high $T > 79$)

- Angry, argumentative, aggressive, impulsive, and nonconforming.
- Generally noncompliant, having difficulty with authority and rules.
- Significant history of antisocial behavior and acting out.

Related Scale Elevations

- Evaluate JCP, ANP, and AGG scales to determine likelihood for acting out against others.
- Substance abuse should be evaluated with the SUB scale.

Treatment Implications

- Evaluate for antisocial personality disorder and even possible psychopathy (though this scale does not measure psychopathy).
- Unlikely to comply with treatment unless externally motivated (mandated) to do so. The therapeutic relationship may be rocky if acting-out behaviors (such as session attendance noncompliance or aggression toward the therapist) are present in treatment.

Ideas of Persecution (RC6)

The RC6 scale evaluates the specific set of beliefs that one is being victimized by others, targeted and controlled by malicious forces. These individuals feel mistreated, and they are suspicious of and have difficulty trusting others. In contrast to RC3, the suspiciousness is a result of feeling specifically targeted, rather than a global feeling that others are only out for their own self-interest. Because of the serious, and rarely endorsed, nature of this scale, its items are considered critical items, and any endorsement should be followed up on.

High Scores on Scale RC6 ($T > 64$; very high $T > 79$)

- Believe others are out to harm, use, or manipulate them.
- Suspicious, mistrustful, and blaming of others, often resulting in withdrawal from or difficulties with interpersonal interaction.
- Generally do not have insight into the potentially problematic thinking.
- Very high scores relate to paranoia and delusional thinking.

Related Scale Elevations

- Evaluate the interpersonal scales (FML, IPP, SAV, SHY, and DSF) for attitudes, values, and behaviors related to RC6.
- Evaluate ANP and AGG to see if ideas of persecution may lead to aggressive behavior toward others.

Treatment Implications

- Assess for paranoid delusions to determine whether antipsychotic medication may be indicated.
- Suspiciousness makes treatment difficult to engage in.

Dysfunctional Negative Emotions (RC7)

The RC7 scale evaluates the presence of a variety of negative emotions, including anxiety, irritability, general unhappiness and helplessness, interpersonal sensitivity, including feeling easily criticized, guilt, and insecurity. Individuals with elevation on RC7 tend to have ruminative thinking patterns. They are worriers and can be extremely reactive to stress and negative interpersonal stimuli. The RC7 scale is highly associated with anxiety disorders.

High Scores on Scale RC7 ($T > 64$; very high $T > 79$)

- Variety of negative emotional experiences, often primarily related to anxiety and worry, but also including anger and self-doubt.
- Thoughts are characterized by rumination, obsessiveness, and uncontrollable intrusive ideation.
- Judge selves harshly and feel guilt and shame easily.
- Highly emotionally reactive to stressful situations and contexts.

Related Scale Elevations

- Evaluate STW, AXY, ANP, BRF, and MSF for specific areas of negative emotion.
- Often highly related to RCd.

Treatment Implications

- Evaluate for anxiety disorders, in order to determine if psychotropic medication may be indicated.
- Discomfort with the emotional distress will likely be a motivating factor for treatment.

Aberrant Experiences (RC8)

The RC8 scale evaluates unusual cognitive, motor, perceptual, and sensory disturbances related to thought disorders. Included may be visual, auditory, or olfactory hallucinations, bizarre sensory/perceptual experiences, and thought broadcasting. High scores suggest impaired ability to test reality and significantly confused thinking, with resultant impaired general functioning. Because of the serious, and rarely endorsed, nature of this scale, its items are considered critical items, and any endorsement should be followed up on.

High Scores on Scale RC8 ($T > 64$; very high $T > 74$)

- Unusual sensory, perceptual, and thought processes, including possible hallucinations and delusions.
- Disorganized and unrealistic thinking, as well as impaired reality testing.
- Possible psychotic disorder.

Related Scale Elevations

- Evaluate RC6 for possible related paranoia.
- Evaluate RC2, SAV, and DSF for complementary symptoms for a schizophrenia diagnosis.

Treatment Implications

- Assess for positive psychotic symptoms to determine whether antipsychotic medication may be indicated.
- Significantly disorganized thinking can deter them from seeking treatment, as well as disrupting the therapeutic alliance.

Hypomanic Activation (RC9)

The RC9 scale evaluates thoughts, feelings, and behaviors related to manic and hypomanic states, including high energy, elevated mood, minimal need for sleep, high self-regard, grandiosity, sensation seeking, risk taking, irritability, poor ability to control impulses, and possible aggression. Elevations may suggest a hypomanic or even manic episode, but moderate elevation may indicate persons who are simply high energy though well adapted.

High Scores on Scale RC9 ($T > 64$; very high $T > 75$)

- High energy and activation, as well as a great deal of engagement with the environment around them and other people.
- Easily become restless and bored, then can become impulsive, aggressive, and highly excitable.
- Irritable, with rapidly shifting mood.
- Euphoria and grandiosity that may characterize narcissistic tendencies.

Related Scale Elevations

- Evaluate AGG, ACT, and SUB to contextualize interpretation of RC9.
- Evaluate RC6 and RC8 for evidence of possible psychotic Bipolar or Schizoaffective Disorder diagnoses.

Treatment Implications

- Mood stabilizing medication may be warranted.
- High activation levels may interfere with treatment, as the slow pace of therapeutic intervention may bore or frustrate them.

MMPI-2-RF SPECIFIC PROBLEM SCALES

Somatic/Cognitive Scales

MLS/Malaise. The MLS scale is a general measure of the extent to which the respondent feels generally physically debilitated. Although it will not specify *how* the individual feels physically weak (the other somatic scales will do this), it measures the general extent to which the person feels less than physically healthy. As is the case with pretty much any self-report measure of somatic complaints, the MLS scale does not distinguish among physical problems related to an actual medical condition, those related only to psychological factors, and those medically related but exacerbated by psychological factors. Information from outside the MMPI-2-RF is necessary to distinguish these. Elevations on MLS ($T > 64$) relate to general feelings of weakness, malaise, fatigue, and general infirmity and poor health. Very high scores ($T > 79$) relate to a greater likelihood of intrusive preoccupation with health concerns, as well as a lack of energy and vitality.

GIC/Gastrointestinal Complaints. The GIC scale measures complaints related specifically to gastrointestinal problems, such as nausea, vomiting, and loss of

appetite. This scale also relates to frequent and recurrent stomachaches, which are typical of many medical and psychological disorders and often at least exacerbated by stress. Elevations on GIC ($T > 64$) very straightforwardly represent the report of gastrointestinal problems, regardless of their etiology.

HPC/Head Pain Complaints. The HPC scale measures complaints related specifically to head pain problems, such as headaches and neck pains. These head pains result in preoccupation with physical problems, and they tend to get worse with stress and negative emotions. The head pains are likely to impair attention and concentration. Elevations on HPC ($T > 64$) very straightforwardly represent the report of head and neck pain problems, regardless of their etiology.

NUC/Neurological Complaints. The NUC scale measures complaints related specifically to neurological and pseudoneurological (neurological-like) problems, such as dizziness, numbness, and motor problems like impaired or involuntary movement. A neurological evaluation may be necessary to evaluate the potential for a genuine medical etiology of the symptoms, though higher elevation on the scale tends to indicate that the symptoms are at least partly psychological in nature, given the broad and diffuse array of symptoms represented on the scale. Elevations on NUC ($T > 64$) very straightforwardly represent the report of vague neurological problems, regardless of their etiology, while very high scores ($T > 91$) represent widespread and global report of neurological problems.

COG/Cognitive Complaints. The COG scale measures complaints related to cognitive functioning, including attention, concentration, memory, and learning. These complaints, much like the somatic ones discussed earlier, may or may not be due to medical or other (e.g., substance abuse) conditions. The COG scale cannot determine the reason for the complaints (or even their genuine presence), but it reflects the degree to which the respondent is reporting them. The COG scale is not a measure of actual cognitive functioning. Elevations on COG ($T > 64$) very straightforwardly represent the report of cognitive difficulties, such as problems with attention, concentration, and memory, regardless of their etiology.

Internalizing Scales

SUI/Suicidal/Death Ideation. The SUI scale assesses thoughts and recent behaviors related to harming oneself and attempting suicide. A single endorsed item on this scale is enough to elevate the T score, and each of the items on SUI is tagged as a critical item and should be evaluated individually and followed up on if endorsed in the problematic direction. Elevations on SUI ($T > 64$) reflect a history of suicidal ideation, intent, or attempts. The higher the elevation, the more likely the suicidal ideation is current. Elevation on this scale is especially critical if there is indication of poor impulse control or behavioral disinhibition, as reflected by elevation on the BXD, RC4, RC9, DISC-r, and SUB scales. As noted, any elevation on the SUI scale should be followed up on immediately, to evaluate risk for suicide and self-harm.

HLP/Helplessness/Hopelessness. The HLP scale assesses beliefs that the individual is incapable of handling any problems he or she is having and that there is little hope for change or improvement in the future. These beliefs can result in a range of outcomes, from belief that the person cannot be helped to an actual internalized lack of

motivation to attempt any change in his or her life. This scale is tagged as critical, so individual items on this scale should be followed up on if endorsed in the problematic direction. Elevations on HLP ($T > 64$) reflect reported feelings of helplessness and hopelessness, with very high scores ($T > 79$) relating to the individual's greater and likely more entrenched beliefs about lack of ability to affect problems and lack of hope that they are likely to improve. Individuals with these elevations are likely overwhelmed with life.

SFD/Self-Doubt. The SFD scale assesses lack of self-esteem and confidence in an individual's own usefulness. Elevations on SFD ($T > 64$) reflect insecurity and feelings of ineffectiveness and inferiority, with very high scores ($T > 69$) relating to feelings of uselessness and self-deprecation. These individuals compare themselves unfavorably to those around them and tend to ruminate on inferior or unimpressive qualities about themselves.

NFC/Inefficacy. The NFC scale assesses the beliefs and feelings that an individual is indecisive, ineffectual, and generally lacking in the ability to accomplish tasks and overcome obstacles in life. Elevations on NFC ($T > 64$) reflect feeling ineffective and indecisive, as well as unable to cope with problems. These individuals are generally passive interpersonally and do not seek out leadership roles. In contrast, low scores on NFC ($T < 39$) reflect a general tendency toward independence, confidence, and leadership roles.

STW/Stress/Worry. The STW scale assesses proneness toward worrying about specific things, such as finances, disappointments, and negative experiences. Elevations on STW ($T > 64$) reflect a tendency to worry about things more than others do, including situations that others would not necessarily worry about. Additionally, these elevations relate to how reactive an individual is to stressors and how much he or she allows those worries to dominate his or her mental space.

AXY/Anxiety. In contrast to STW, which focuses on ideational worry about specific situations and circumstances, the AXY scale assesses the global experience of more pervasive anxiety. Included on this scale are symptoms like nightmares, constant anxiety and fear of impending negative events, and worrying every day. These anxieties may be related to posttraumatic exposure (as is the case in Posttraumatic Stress Disorder), though they may not necessarily be. The items on this scale are considered critical, and each one should be evaluated individually and followed up on immediately if endorsed in the problematic direction. Elevations on AXY ($T > 64$) reflect constant, pervasive, intrusive anxiety over which the individual has little or no control. Such individuals likely have ruminative and intrusive thoughts and difficulties sleeping. Elevations should trigger the hypothesis of a possible anxiety-related disorder.

ANP/Anger Proneness. The ANP scale assesses shortness of temper and tendency to be impatient with others and quick to anger. As one of the internalizing scales, ANP relates more to the internal feelings of anger rather than the external behaviors of aggressiveness (which are represented on the AGG scale). Elevations on ANP ($T > 64$) relate to being easily upset and quick to anger. With higher elevations, individuals may become overwhelmed by anger and feel that it is uncontrollable. These individuals tend to be obstreperous and irritable, with poor frustration tolerance.

BRF/Behavior-Restricting Fears. The BRF scale assesses fears and worries that inhibit or alter the normal, daily activities a person would engage in. Included are

fears of open spaces, the dark, leaving home, dirt, and touching money. Each of these fears could easily interfere with everyday activities, in ways that range from minor (such as needing night lights for fear of the dark) to major (such as agoraphobia for fears of leaving home and open spaces). Elevations on BRF ($T > 64$) reflect multiple fears that inhibit or alter behavior significantly. Higher elevations may be related to agoraphobia, which should be evaluated.

MSF/Multiple Specific Fears. The MSF scale assesses fears that are specific to animals and other natural elements and situations. Included are blood, fire, thunder, spiders, and other animals. These fears are consistent with specific phobias, though even endorsement of individual items may not constitute fear to the level of this diagnosis. Elevations on MSF ($T > 64$) reflect the self-report of multiple specific fears. Individuals with these elevations are likely to be somewhat behaviorally inhibited and avoid risky situations. Follow-up should include assessment of possible specific phobias.

Externalizing Scales

JCP/Juvenile Conduct Problems. The JCP scale assesses specific problems related to RC4, including a history of problematic behavior in school, adolescent problem behaviors like stealing, and being negatively influenced by peers. Although the content of JCP focuses specifically on past problems, its relation to other scales, such as BXD, RC4, and AGG, may complete a developmental picture of problematic externalizing problems. Elevations on JCP ($T > 64$) reflect the self-report of a history of behavioral problems, primarily during adolescence. Individuals with these elevations are likely to be oppositional, especially with authority figures, and have difficulties in interpersonal functioning more broadly. Follow-up should include assessment of possible Antisocial Personality Disorder.

SUB/Substance Abuse. Also a specific facet of RC4, the SUB scale assesses the acknowledgement of past and current substance abuse (especially alcohol abuse) and problems resulting from or related to it. This scale is tagged as critical, so individual items on this scale should be followed up on if endorsed in the problematic direction, as they may relate to current problematic substance abuse. As the items on SUB are face valid, it is easy for respondents to deny substance abuse, even if it is a problem. As such, information from outside of the MMPI-2-RF should be evaluated to cross-validate this scale, if it is not elevated. Elevations on SUB ($T > 64$) reflect a significant acknowledged history and/or current problem with abusing alcohol and potentially other drugs. If further evaluation of elevated scores on SUB reveals significant current drug or alcohol abuse, clinicians may consider focusing treatment specifically on this first, even before attending therapeutically to other problems.

AGG/Aggression. The AGG scale assesses specific problems related to RC9, including a reported history of and current problems with physically aggressive behavior. Although likely related to ANP, this scale focuses primarily on the behavioral aspects of anger and aggression rather than on the affective or cognitive experience. This scale is tagged as critical, so individual items on this scale should be followed up on if endorsed in the problematic direction, as they may relate to current proneness toward aggressive behavior. Elevations on AGG ($T > 64$) reflect the self-report of a history of and/or current problems with physical aggression, including intimidation, violent behavior, and

actual physical fights. These individuals likely have anger management problems and may lash out physically toward others in their lives. The potential for violence should be taken seriously within the therapeutic situation.

ACT/Activation. Also a specific facet of RC9, the ACT scale assesses high levels of excitation and energy. Included on this scale are symptoms like not needing sleep, significant mood swings, and noticeably heightened energy levels. Elevations on ACT ($T > 64$) reflect excessive excitation and energy, potentially consistent with a hypomanic or manic episode. Elevations should always be considered within the context of the possibility that activation may be substance induced.

Interpersonal Scales

FML/Family Problems. The FML scale assesses negative family situations. These situations may include a generally tumultuous family environment, not feeling that the person is able to rely on his or her family in times of need, and generally negative feelings within the family unit, including not feeling appreciated or valued. Elevations on FML ($T > 64$) reflect significant family turmoil and negative feelings. The nature of these family problems should be followed up on to see how they may relate (lead to, result from, etc.) with any other elevations.

IPP/Interpersonal Passivity. The IPP scale assesses the bidirectional trait of assertiveness and passivity. Items on the scale are keyed to reflect passivity, including not wanting to lead or take control of situations, acquiescing to others' wishes easily, and being generally interpersonally submissive. Elevations on IPP ($T > 64$) reflect a general predisposition toward submissiveness and passivity, including not wanting to create conflict by disagreeing with others. Extreme elevations may be related to pathological submissiveness, such as in Dependent Personality Disorder. Low scores on IPP ($T < 39$) reflect assertiveness and strong leadership qualities, being unafraid to stand up for him- or herself and make his or her point of view heard and known. Extremely low scores may be related to narcissism. As this scale is predominantly behavioral, the *reasons* for the passivity can be varied and may be related to elevations on other scales.

SAV/Social Avoidance. The SAV scale assesses the bidirectional level of liking or disliking social situations. Items on the scale are keyed to reflect lack of enjoyment and general avoidance of social situations. Elevations on SAV ($T > 64$) reflect a dislike and avoidance of situations that are primarily social in nature. Extreme elevations may be related to disorders that include avoidance, such as Avoidant Personality Disorder, Social Phobia, or Schizoid Personality Disorder. Notably, elevations on SAV without elevations on SHY (see the next section) likely relate to social avoidance that is *not* due to anxiety. Low scores on SAV ($T < 39$) reflect gregariousness, a general comfort in and enjoyment of social situations, and a generally outgoing personality.

SHY/Shyness. The SHY scale assesses discomfort with social situations and is often related as a precursor to social avoidance. Items on this scale relate to feeling easily embarrassed or scrutinized by others, general shyness, and feeling uncomfortable in social situations. Elevations on SHY ($T > 64$) reflect significant anxiety about socializing with others. This anxiety may be related to a diagnosis of Social Phobia or Avoidant Personality Disorder, especially if behaviorally related to elevations in SAV.

DSF/Disaffiliativeness. The DSF scale assesses general asocial tendencies and, along with SHY, may be an underlying factor related to social avoidance. Items on this scale relate to a dislike of interacting with others; a lack of close, intimate relationships; and preferring to be on one's own. Elevations on DSF ($T > 64$) reflect asocial tendencies, preferring to be alone and not having significant, close relationships. Likely difficult to engage in a therapeutic relationship, extreme elevations may be related to Schizoid Personality Disorder.

INTEREST SCALES

Aesthetic-Literary Interests (AES)

The AES scale reflects general interest in activities related to aesthetics and artistic expression, such as writing, music, dance, and theater. Elevations on AES ($T > 64$) reflect significant interest in these types of activities. Individuals with these interests tend to be interpersonally empathic and highly attuned to their own sensory experience.

Mechanical-Physical Interests (MEC)

The MEC scale reflects general interest in activities related to physical activity and mechanics, such as sports, outdoor activities like camping, and fixing things. Elevations on MEC ($T > 64$) reflect significant interest in these types of activities. Individuals with these interests tend to like adventure and excitement.

MMPI-2-RF PERSONALITY PSYCHOPATHOLOGY FIVE SCALES

The Personality Psychopathology Five (PSY-5) scales represent a special cluster of scales, core to the MMPI-2-RF but also usable on the MMPI-2 (Harkness et al., 2002). Item selection was preceded by working with a group of laypersons to develop relevant, distinctive, and easily understandable personality constructs. The five emerging constructs were then used to select existing MMPI-2 items reflecting these constructs. The resulting items were refined by professional reviewers and then submitted to formal psychometric analysis. The descriptors summarized next are from the Harkness et al. (2002) monograph, Ben-Porath (2012), and the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008/2011).

Aggressiveness–Revised (AGGR-r)

The AGGR-r scale assesses general personality traits related to interpersonal passivity, assertiveness, and aggressiveness. These concepts relate to both the tendency to stand up for oneself *and* the level of respect with which one does so (Lange & Jakubowski, 1976). That is, both assertiveness and aggressiveness include standing up for oneself, but the former does so respectfully, while the latter involves infringing on another's rights. Passivity involves respect and not standing up for oneself, while

passive aggressiveness involves not standing up for oneself but also not interacting in a respectful way. This scale, however, treats these three behavioral, interactive tendencies (passivity, assertiveness, and aggressiveness) along a single continuum, with low scores relating to passivity, moderate scores relating to assertiveness, and high scores relating to aggressiveness.

High Scores on AGGR-r ($T > 64$)

- Enjoys intimidating others, being overly socially dominant.
- Aggression used to accomplish goals.
- Dominant and extroverted.
- Belief in own leadership abilities.
- Possible history of being antisocial and physically abusive.
- High-scoring men are more likely to have a history of domestic violence.
- High-scoring women are more likely to have been arrested.
- Often related to low IPP, as well as high ANP and AGG.

Low Scores on AGGR-r ($T < 39$)

- Interpersonally passive, submissive, and potentially dependent.
- Overly acquiescent to the wishes of others.

Diagnostic and Treatment Implications

- High scores may relate to Cluster B personality disorders.
- The therapeutic relationship will likely thrive with a balancing of this trait, such that elevations should be met with a more supportive therapist and low scores should be met with a more directive therapist.

Psychoticism–Revised (PSYC-r)

The PSYC-r scale does not relate closely with any scale on the five-factor model of personality, suggesting that it may be more useful as it relates to personality psychopathology than to normal personality functioning. The scale assesses abnormalities in sensation, perception, and thinking.

High Scores on PSYC-r ($T > 64$)

- Unusual thought processes, including disorganized thinking, impaired reality testing, and tangential, bizarre, or disoriented thoughts.
- Unusual thought content, including delusions of reference.
- Unusual sensory or perceptual phenomena, including hallucinations.
- Alienated from others.
- Outpatients were described as having low functioning, few or no friends, and depressed.
- High-scoring men were rated as sad and depressed.
- High-scoring women were more likely to be experiencing hallucinations.

- Inpatient populations are more likely to be psychotic and have paranoid suspiciousness, loose associations, flights of ideas, hallucinations, and ideas of reference.
- Often related to elevations on THD.

Diagnostic and Treatment Implications

- High scores may relate to Cluster A personality disorders.
- Distortions in thinking and reality testing can make developing a trusting therapeutic relationship difficult.

Disconstraint–Revised (DISC-r)

The DISC-r scale assesses a variety of behaviors related to “disconstraint,” or not keeping control over one’s behaviors. Included on this scale are impulse control, excitement-seeking behaviors, and a tendency to act out behaviorally.

High Scores on DISC-r ($T > 64$)

- Engages in impulsive and acting-out behaviors.
- Risk taking, antisocial, aggressive, with a possible history of being arrested.
- Nontraditional and easily bored.
- Possible history of having abused drugs and alcohol.
- High-scoring men have more histories of domestic violence.
- High-scoring women are somewhat achievement-oriented.
- Often related to RC7.

Low Scores on DISC-r ($T < 39$)

- Extremely strict self-control and reduced impulsivity, behaviorally constrained.
- Adhere closely to rules.
- Can easily tolerate boredom, and prefer romantic partners who are also constrained.

Diagnostic and Treatment Implications

- High scores may relate to Cluster B personality disorders.
- High scorers may have difficulty with treatment, as they are likely not to be internally motivated to comply and can impulsively terminate treatment.

Negative Emotionality/Neuroticism–Revised (NEGE-r)

The NEGE-r scale assesses the general presence and level of negative emotional experiences. Many of the negative emotional experiences relate to anxiety and depression, including worry, insecurity, and pessimism, and elevations can inhibit behavior in many situations.

High Scores on NEGE-r ($T > 64$)

- High experience of generally negative emotions.

- Worry, guilt, insecurity, self-criticism, and thinking in terms of worst-case scenarios.
- Outpatients are likely to be depressed or dysthymic, with few or no friends; anxious; or have somatic symptoms.
- High-scoring men are likely to have engaged in domestic violence related to attempts at maintaining a focus on the flaws and irritations with their spouse and their future.
- High-scoring women are likely to be pessimistic, have low achievement, and have histories of alcohol abuse.

Diagnostic and Treatment Implications

- High scores may relate to Cluster C personality disorders.
- The subjective experience of negative emotions may serve to motivate them for treatment.

Introversion/Low Positive Emotionality–Revised (INTR-r)

The INTR-r scale assesses the general lack of positive emotional experience and lack of engagement in socialization. Included are both aspects related to low enjoyment in activities (anhedonia) and pessimism, as well as aspects related to social avoidance and withdrawal.

High Scores on INTR-r ($T > 64$)

- Lacks positive emotional experiences and can feel sad, depressed, pessimistic, anxious, or flat.
- Introverted and socially avoidant.
- Low achievement orientation.
- May have somatic symptoms.
- High-scoring women are more likely to have taken antidepressants and to report having few or no friends.
- Often related to RC2.

Low Scores on INTR-r ($T < 39$)

- Energetic, with many positive emotional experiences.
- Good capacity to experience pleasure and joy, unlikely to be depressed or dysthymic.
- Outgoing and social.
- Extremely low scores may suggest hypomanic features.

Diagnostic and Treatment Implications

- High scores may relate to Cluster C personality disorders.
- High scorers should be evaluated specifically for mood disorders, such as dysthymic and depressive disorders.
- Lack of positive emotionality and social engagement may interfere with the building of a therapeutic alliance.

RECOMMENDED READING

- Archer, R. A. (2005). *MMPI-A: Assessing adolescent psychopathology* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis: University of Minnesota Press.
- Butcher, J. N. (2011). *A beginner's guide to the MMPI-2* (3rd ed.). Washington, DC: American Psychological Association.
- Butcher, J. N., & Perry, J. (2008). *Psychological assessment in treatment planning: Use of the MMPI-2 and BTPI*. New York, NY: Oxford University Press.
- Caldwell, A. (2001). What do the MMPI scales fundamentally measure? Some hypotheses. *Journal of Personality Assessment*, 76, 1–17.
- Friedman, A. F., Bolinsky, P. K., Levak, R. W., & Nichols, D. S. (2014). *Psychological assessment with the MMPI-2/MMPI-2-RF* (3rd ed.). New York, NY: Routledge/Taylor & Francis.
- Graham, J. R. (2011). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.
- Levak, R. W., Siegel, L., & Nichols, D. S. (2011). *Therapeutic feedback with the MMPI-2: A positive psychology approach*. New York, NY: Routledge.

PERSONALITY ASSESSMENT INVENTORY

The Personality Assessment Inventory (PAI) is a self-administered, paper-and-pencil/online test composed of 344 statements for which the respondent must choose how true each is for him or her. Each item is rated on a four-alternative scale, including *False, Not at all True (F)*; *Slightly True (ST)*; *Mainly True (MT)*; and *Very True (VT)*. The test can be administered either to individuals or groups and was developed to assess adults between the ages of 18 and 89. There is also an adolescent version, the PAI–A, for use with adolescents aged 12 to 18. The PAI requires only a fourth-grade reading ability. The PAI items request information concerning an individual's behavior patterns, current feelings and opinions, and reactions to different situations. The results are plotted on 22 nonoverlapping scales of four different kinds: 4 validity scales, 11 clinical scales, 5 treatment-related scales, and 2 interpersonal scales (see Table 8.1).

The conceptual orientation of the PAI is based on a balance of rational and quantitative methods of scale development. The selection of syndromes to assess was originally based on how stable their importance was in the history of understanding mental illness, as well as their importance in current diagnostic practice. Items for each scale were developed based on extensive reviews of both historical, conceptual literature and contemporary, empirical literature, focusing on the concepts that are central and core to the concepts of each construct. Rather than a purely empirical approach, such as criterion-keying or a factor-analytic approach to item selection, this theoretical grounding allows each scale to include a full range of severity of each construct, including both milder and more severe indicators. Balancing this rational, theoretical approach to item development has been rigorous empirical scrutiny on the items and scales.

The PAI was originally developed by Leslie Morey and published in its original form in 1991. The reviews of the test have generally been significantly positive, with the exception of some criticism over some scales (particularly related to personality disorders) potentially overlooked for inclusion. The key to interpreting scales is that each scale was designed to assess a particular construct, represented by the name of the scale. This in some ways makes interpreting the PAI much more straightforward than interpreting the Minnesota Multiphasic Personality Inventory (MMPI; in which scales are more related to general personality traits than diagnostic categories) and even the Millon Clinical Multiaxial Inventory (MCMI; in which scales represent a spectrum of personality features, from normal to exaggerated, often adaptive to maladaptive). Additionally, the nonoverlapping nature of the scales, while perhaps not as representative of actual behavioral, personality, or symptomatic traits, allows practitioners to feel

Table 8.1 Validity, Clinical, Treatment, and Interpersonal Personality Assessment Inventory Scales

Name	Abbreviation	No. of items	Mean alpha*
Validity scales			
Inconsistency	ICN	10 pairs	.31
Infrequency	INF	8	.38
Negative Impression	NIM	9	.70
Positive Impression	PIM	9	.74
Clinical scales			
Somatic Complaints	SOM	24	.88
Anxiety	ANX	24	.91
Anxiety-Related Disorders	ARD	24	.81
Depression	DEP	24	.89
Mania	MAN	24	.82
Paranoia	PAR	24	.87
Schizophrenia	SCZ	24	.84
Borderline Features	BOR	24	.88
Antisocial Features	ANT	24	.85
Alcohol Problems	ALC	12	.87
Drug Problems	DRG	12	.76
Treatment scales			
Aggression	AGG	18	.88
Suicidal Ideation	SUI	12	.88
Stress	STR	8	.75
Nonsupport	NON	8	.76
Treatment Rejection	RXR	8	.76
Interpersonal scales			
Dominance	DOM	12	.80
Warmth	WRM	12	.81

Source: Adapted from *Personality Assessment Inventory Professional Manual*, 2nd ed., by L. C. Morey, 2007, Odessa, FL: Psychological Assessment Resources.

*Mean alpha was calculated from reported alphas from studies of three separate samples, including a census sample, a college sample, and a clinical sample.

comfortable that elevation on one scale is not causing erroneous elevation on another. Also, the four-alternative choice format allows the test to add a deeper dimension of severity to the findings, such that clients can elevate a scale slightly with an endorsement of *Slightly True* to a symptom description, but they can elevate the scale more significantly with a *Very True* endorsement, representing the likelihood that the individual is exhibiting that symptom more significantly or constantly.

HISTORY AND DEVELOPMENT

One of the newer tests to gain widespread favor in the field, the PAI was originally developed between 1987 and 1991, with its original version being published in 1991. It was originally developed to be a measure that holds strong construct validity and also provides clinically useful information, very much aligned with the nosology of the revised third edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R*; American Psychiatric Association, 1987). The original scales and subscales were selected based on systematic reviews of the literature, surveys of test users, and the incorporation of the diagnostic schemas (namely the *DSM-III-R*) that were widely used at the time. An original pool of 2,200 items was evaluated first by four experts, who were tasked with matching individual items to the content-driven scales. Any item with at least 75% agreement was retained. The item pool was then evaluated by a bias panel to check if items could be considered offensive or exclusionary for gender, race, ethnic, or religious groups, and such items were eliminated. Finally, external groups of experts in the field were tasked to match individual items to scales, similar to the original group of four, and high overall agreement resulted in 776 items.

This 776-item alpha version was administered to college students in three different conditions: a standard, “regular” administration, one in which the students were asked to present themselves in a favorable way, and one in which the students were asked to fake symptoms of mental disorders. Analysis of the results of these trials led to reduction to a 597-item beta version, based on response variability, correlation with other items on the scale, low correlation with other scales, absence of gender differences, and low correlation with measures of social desirability and both positive and negative self-presentation. A similar process was conducted again on this beta version, except instead of using college students, a community sample, a clinical sample, and college students either presenting favorably or malingering mental illness were tested. Based on similar criteria as when they reduced the alpha version, the final scale included 344 items, which were used for norming. Although the test manual was revised in 2007, there have been no major revisions to the test since its first publication.

RELIABILITY AND VALIDITY

In general, the reliability of the PAI compares favorably with other personality inventories. Morey (2007) reported reliability analyses from the standardization sample. Internal consistency (as measured by alpha coefficients) for individual scales (excluding the validity scales) ranged between a low of .75 for Stress to a high of .91 for Anxiety (see Table 8.1), with the median alpha being .845. Subscales also had adequate reliabilities. Alpha coefficients for the subscales ranged between a low of .54 for Activity Level (MAN-A) to a high of .85 for Traumatic Stress (ARD-T). The median alpha for the subscales was .74. Others have reported on internal consistency of the index scores for different populations, including low socioeconomic status, ethnic minority patients on methadone (Alterman et al., 1995), outpatients with substance abuse (Rogers, Flores,

Ustad, & Sewell, 1995), psychiatric inpatients (Boone, 1998; Siefert, Sinclair, Kehl-Fie, & Blais, 2009), women with eating disorders (Tasca, Wood, Demidenko, & Bissada, 2002), and candidates for bariatric surgery (Corsica, Azarbad, McGill, Wool, & Hood, 2010). Generally, internal consistency of the scales has been found to be relatively consistent with the adequate to excellent outcomes with the standardization sample.

From the standardization sample, test-retest reliabilities for individual scales (excluding the validity scales) on the PAI ranged between a low of .68 for Dominance to a high of .92 for Alcohol Problems, with the median test-retest reliability being .83. Test-retest reliabilities for the subscales ranged between a low of .68 for Conversion (SOM-C) and Activity Level (MAN-A) to a high of .85 for Cognitive Anxiety (ANX-C), Affective Depression (DEP-A), and Stimulus Seeking (ANT-S). The median test-retest reliability for the subscales is .78. Morey (2007) also reported on code type stability, examining how consistent the single highest clinical scale was upon readministration of the PAI. When examining the full sample, 57.4% of the individuals had the same highest scale on both administrations of the test. When only examining those protocols with significant elevation, meaning at least one scale with a *T* score above 70, the number of individuals with the same highest scale rose to 76.9%. These numbers demonstrate relatively good stability of the clinical scales, especially when they are significantly elevated. It is important to note that these analyses did not incorporate analysis of 2-point code types or other complex configurations.

Factor-analytic studies have consistently reported, for both community and clinical samples, a general four-model structure for the PAI (Deisinger, 1995; Morey, 2007; Schinka, 1995). Three of the four factors include a general negative adjustment factor, onto which the Depression, Anxiety, Anxiety-Related Disorders, Schizophrenia, and Borderline Features scales load significantly; a behavioral acting-out factor, onto which Antisocial Features, Alcohol Problems, and Drug Problems significantly load; and a factor relating to narcissism and interpersonal exploitation, onto which the Mania, Antisocial Features, and Dominance scales significantly load. The fourth factor emerged differently for the clinical and community samples, with the former representing a carelessness factor (defined by the Infrequency and Inconsistency scales) and the latter representing a social detachment factor (defined by low Warmth and high Nonsupport, Schizophrenia, and Paranoia scales). It is important to note that at least 25% of the variance in the entire measure is not accounted for in the four-factor model, suggesting that there is clinical utility in using the individual scales separately, rather than sacrificing them for a four-factor interpretation.

The PAI manual (Morey, 2007) includes a great deal of information related to the PAI scales and subscales as they correlate with other measures of personality and psychopathology. Among *many* other measures, the PAI scales and subscales were compared to appropriate scales from the MMPI, NEO, MCMI, Beck Anxiety Inventory, Beck Depression Inventory, Brief Psychiatric Rating Scale, Cognitive Distortion Scale, Inventory for Altered Self-Capacities, California Psychological Inventory, Buss-Perry Aggression Questionnaire, and even the Rorschach. Those interested are encouraged to read Chapter 9 in the PAI manual for many details. In general, Morey builds a very strong case for each scale and subscale, with moderate to good correlations generally observed for most. An important feature of these presented findings is the inclusion of many different types of samples, including community,

clinical (inpatient and outpatient), forensic, and student, among others. Each scale is presented with a variety of samples and comparison measures by which it is validated.

In addition to these criterion-based validity studies, the manual (Morey, 2007) also presents research related to scales and subscales differentiating normal respondents and the scales' intended clinical samples. For example, researchers evaluated whether or not the Depression scale could differentiate respondents with and without depression. A great deal of data is presented in the manual, systematically following each of the scales and reviewing literature. Again, nearly all the scales do an adequate job of differentiating clinical groups. However, there are some exceptions. Most notably, the findings on the ability of the Schizophrenia (SCZ) scale to distinguish individuals with and without psychosis has been mixed (Edens, Cruise, & Buffington-Vollum, 2001; Rogers, Ustad, & Salekin, 1998). Also, as is noted later in the chapter in the sections titled "Positive Impression (PIM, DEF, CDF, ALC Est, DRG Est)," "Alcohol Problems," and "Drug Problems," the Alcohol Problems (ALC) and Drug Problems (DRG) scales ask direct questions about substance use behaviors and their consequences. As such, they are highly susceptible to denial or motivated underreporting (Edens et al., 2001; Fals-Stewart, 1996). Their resultant discriminative ability is not as strong as many other scales. However, the use of ALC Est and DRG Est calculations (discussed later in the "Positive Impression" section) can help improve the predictive power of the PAI with populations who abuse drugs and alcohol.

ASSETS AND LIMITATIONS

The PAI presents some distinct advantages over other self-report measures of personality and psychopathology. First and foremost, its development focused on a balance between psychometric adequacy, linkage to the *DSM* as the most common diagnostic tool in the field currently, and clinical utility. While the overall scales maintain excellent reliability and adequate validity, the subscales especially add a clinically useful dimension to the measure. Creating these subscales based on clinical utility rather than factor analysis has some drawbacks. For example, some research with the Borderline Features subscales has found a six-factor structure to be a better fit than the four factors represented by the current subscales (K. M. Jackson & Trull, 2001). However, what may be lost in the psychometrics of these subscales is made up for in the clarity by which they delineate differential aspects of different disorders. For example, the breaking down of general anxiety and depression into components related to cognitive, affective, and physiological correlates of the syndromes can help clinicians understand how certain interventions may better help clients. While these subscales (perhaps with the exception of the Anxiety-Related Disorders scale, on which the subscales represent different types of syndromes or disorders) are not meant to diagnose, they can add nuance to the understanding of what is going on with the examinee.

From the perspective of the test-taker, one major asset of the PAI is its low reading level, which is reported as equal to fourth-grade reading ability. The PAI is often also praised for its relatively straightforward wording of items, especially as compared to the MMPI. The low required reading ability level and the brevity of items makes the instrument relatively quick to take, with reported averages less than an hour.

The four-alternative response format also allows respondents the ability to provide some important dimensional information on items, rather than being forced to choose the absolutes of true and false. This four-alternative scale format, in addition to offering examinees the potential for responses to be closer to their own experience, also adds variability to the PAI that allows not only for information about whether problems may be present, but also information about severity and depth. That is, elevations on scales can relate to moderate ratings on many of the scales' items or severe ratings on a few. Analysis of subscales can also help with these distinctions.

An additional asset of the PAI is related to the potential for multiple or alternate comparisons in order to best meet the needs of the particular case. The original PAI standardization was completed on three different samples representing three different populations. Included were a community sample, a clinical sample, and a college sample. The typical presentation of scores plotted on the profile form shows the respondent in comparison to the general, census-matched community sample (the primary *T* score), as well as in comparison to the clinical sample, represented by a "skyline" presentation of scores 2 standard deviations above the mean for the clinical sample. Appendices in the professional manual present data to compare examinees' scores to the census-matched community sample, the clinical sample, a census-matched African American community sample, a census-matched sample of older adults (60 years and older), and college students, adding to the interpretive flexibility of the test.

Another major asset of the PAI is its nosological alignment with the *DSM*. The scales generally align clearly with specific disorders listed in the diagnostic manual. However, this alignment is related to one of the PAI's greatest limitations. While no scale can encompass the entire range of potential mental illness, the PAI is saliently missing several scales that would relate directly to *DSM* disorders, necessitating the supplementing of the PAI with other measures aimed directly at these problems. Perhaps most notably missing is information related to eating disorders. However, a major criticism has been related to the lack of scales relating to the majority of personality disorders; the PAI only has named scales for Borderline and Antisocial Features. It should be noted that there are configurations of scales and subscales that inform diagnosis of other personality disorders. For example, information about dependent personality disorder can be obtained from high scores on Warmth (WRM) and low scores on Dominance (DOM), Verbal Aggression (AGG-V), and Grandiosity (MAN-G). Additionally, it should be noted that some have argued that a single, general factor of personality accounts for much of the variance of personality disorders as assessed by self-report inventories, including the PAI (Rushton & Irwing, 2009). As such, developing scales for other personality disorders that include items that do not overlap with those already on the PAI would be difficult to accomplish and potentially psychometrically problematic, even if clinicians might find such scales clinically useful.

Finally, also related to ease of applicability of the PAI scales and subscales to diagnosis (*DSM* or otherwise), a limitation of the measure is found on its Alcohol Problems (ALC) and Drug Problems (DRG) scales. Beyond the validity questions that have already been raised about these scales, there is a limit to their clinical utility because the items on these scales encompass both current and past substance use. Elevated scores on either of these scales are difficult to interpret because of this structure. While many of the items are worded in the past tense, the distinction between past and present use

of alcohol and drugs is extremely important in the evaluation and diagnosis of substance use problems. It is important to note that elevations on either of these scales (or elevations on ALC Est or DRG Est, described later) should trigger further evaluation.

USE WITH DIVERSE GROUPS

Research on the PAI with different racial and ethnic groups in the United States has generally been positive. African Americans are well represented within the normative sample. One major criticism of the measure has been in its use with Latino/a populations, which are inadequately represented in the normative sample (which is based on census-matched data that includes race—Black, White, and other—but not Latino or Hispanic ethnicity; Alamilla & Wojcik, 2013). However, use with this population suggests mostly reliable, valid, and useful results (Hopwood, Flato, Ambwani, Garland, & Morey, 2009), though a simplified internalizing/externalizing structure may fit better with Latino/a Americans (Hopwood & Moser, 2011). Research with samples from low socioeconomic backgrounds has shown similar psychometric properties and utility to the normative sample, but with worse internal consistency on many scales (Alterman et al., 1995; Donaldson, 2010).

The Spanish version of the PAI initially exhibited some problems with its internal consistency when evaluating a Mexican American sample (Rogers, Flores, Ustad, & Sewell, 1995). However, later research revealed that performance across the scales is similar to the English version (Fernandez, Boccaccini, & Noland, 2008) and that results are not significantly influenced by ethnic background (Fantoni-Salvador & Rogers, 1997). The German version of the test also has a similar factor structure as the English version (Groves & Engel, 2007). Similarly, the Chinese-language version of the PAI had a similar factor structure to the English version (Cheung, Cheung, Leung, Ward, & Leong, 2003), similar gender differences to the English version (Cheung, Leung, Fan, Song, Zhang, & Zhang, 1996), and similar scale functioning (Lin & Church, 2004). However, some researchers have argued that some personality concepts that are core to Chinese culture, such as graciousness, family orientation, inner and interpersonal harmony orientation, and attitudes toward traditional Chinese beliefs, are not reflected in the measure (Cheung et al., 1996).

The sample of research just discussed provides moderately good support for using the PAI in cross-cultural contexts. However, additional research still needs to be conducted on the relationship between PAI scores and race, socioeconomic status, and other demographic variables. In particular, further research needs to be conducted on the ability of the PAI to predict relevant behaviors in a specific cultural context. Often, knowing that there are differences in various scale scores is not sufficient. We also need to know the meanings of these differences for understanding and predicting relevant external behaviors and internal experiences.

INTERPRETATION PROCEDURE

Effective interpretation of the PAI is relatively straightforward, as each scale is designed to include content directly related to the name of that scale. That is, an individual who

is significantly elevated on the Anxiety (ANX) scale is very likely to be reporting experienced symptoms of anxiety. However, the subscales, frequent 2-point code types, and other configurations can certainly add nuance to the direct interpretation of individual scales. For example, two individuals may score similarly on the Depression (DEP) scale; however, an analysis of the subscales that comprise the DEP scale may reveal that one has significant elevation on the Cognitive Depression (DEP-C) subscale while the other has much more significant elevation on the Physiological Depression (DEP-P) subscale. The former individual's possible depression would be more related to thoughts of helplessness, hopelessness, and worthlessness, whereas the latter individual's possible depression is more centered on decreased activity, energy, and changes in sleep and appetite. Similarly, two individuals may be elevated in the same way on the Anxiety (ANX) scale, but one is also elevated on Depression (DEP) while the other is secondarily elevated on Paranoia (PAR). These individuals' anxiety would look quite different, with the former more moody and tense and the latter more suspicious and hostile.

Further, an individual who presents with an elevated Borderline Features (BOR) scale, and especially if all four subscales are elevated, may have borderline personality disorder. However, other scales may also contribute evidence of borderline personality disorder, such as Stress (STR), Traumatic Stress (ARD-T), and Suicidal Ideation (SUI). Similarly, how an individual feels about him- or herself can be informed by several scales evaluated concurrently, such as Cognitive Depression (DEP-C), which includes components of self-efficacy, Grandiosity (MAN-G), which includes components of self-esteem, and Identity Problems (BOR-I), which includes components of how stable one's self-concept is. Evaluating scales together can offer a more nuanced and informative description of the experience of the client. Interpretation can be organized according to the next five steps.

1. Interpreting Test Validity

As with many self-report personality measures, the first step in interpretation of the PAI is an evaluation of the profile validity, which is a process that includes evaluation of the validity scales as well as evaluation of other factors that may produce potential distortion. Situational factors, such as potential for secondary gain (e.g., a client claiming emotional damage in a lawsuit, parents being evaluated in a child custody evaluation), and personal factors, such as limited insight or personal shame (e.g., about the stigma of mental illness), have the potential to distort individuals' responses. A clinical evaluation of the circumstances and potential for response distortion should be considered in the assessment of profile validity.

Each of the validity scales, then, provides more detailed information related to profile validity. The PAI does not use the validity or any other scales to modify any scale scores. The practitioner must evaluate the validity scales and make a decision about whether and how to proceed to interpreting the other scales. Two of the validity scales, the Inconsistency (ICN) and Infrequency (INF) Scales, are designed to detect idiosyncratic responding. The other two, Negative Impression (NIM) and Positive Impression (PIM), are designed to detect response bias aimed at managing impressions on the measure. Elevations on any of these four validity scales should signal that, at the very least, the profile should be interpreted with caution, if not discarded entirely. In addition

to these primary validity scales, Morey (1996) has provided several configural scales, including the Defensiveness Index and the Malingering Index. Each of these takes into account various scales and subscales that, taken together, are typically only present in purposefully distorted profiles.

2. Analysis of Critical Items

Once profile validity has been evaluated, four additional levels of interpretation can be undertaken for the PAI Clinical, Treatment, and Interpersonal scales. First, individual items can be interpreted on the PAI, much more so than on other measures, like the MMPI or MCMI, because each item was developed due to its centrality to the overall construct of the scale it is on. Twenty-seven items are classified as “critical items,” because of the relative infrequency of their endorsement by “normal” individuals and because of their potential relevance to crisis situations.

3. Full Scale Interpretation

The next level of interpretation for the PAI scales is the full scale level. Each full scale score is compared to two different normative groups, the general population and a clinical sample. The *T* scores provided by the PAI are directly interpretable as a comparison to a representative community sample. The profile form also provides a profile “skyline,” which represents the *T* scores on each scale that are 2 standard deviations above the mean for a clinical sample of adults. Individuals who score above this skyline on any scale are elevated even in comparison to patients in a clinical setting. The professional manual (Morey, 2007) also includes means and standard deviations necessary to convert raw scores to *T* scores for comparison with multiple populations, including college students, a clinical sample, older adults, and others. As stated previously, the interpretation of full scale scores is relatively straightforward, with each scale constructed to rationally represent the construct of the scale name. Elevation on the Somatic Complaints scale (SOM), for example, is related directly to a likelihood of the individual struggling with health concerns and physical problems like those seen in somatization and conversion disorders.

4. Subscale Interpretation

Following the interpretation of full scale scores, practitioners can use the subscale scores to further understand, clarify, and define the components of the full scales. The subscales were developed conceptually (rather than purely empirically) in order to cluster rationally similar items in clinically relevant ways. For example, elevation on the Anxiety (ANX) full scale may indicate the presence of clinically significant anxiety or even an anxiety disorder. However, scrutiny of the subscales that comprise ANX may reveal a different reason for the elevation. Many individuals who come for an evaluation are in some sort of distress, anxious about their own suffering or often anxious about the testing situation itself. Individuals with an elevation on the Cognitive Anxiety subscale (ANX-C), which focuses on worry, in the absence of elevation on the Affective Anxiety (ANX-A) and Physiological Anxiety (ANX-P) subscales, which

focus on the emotional and physical signs of tension and stress, respectively, are more likely worried about the situation or their own suffering, rather than suffering from an anxiety disorder. Individuals who are elevated on all three ANX subscales, however, are more likely to be suffering from an actual anxiety disorder. A review of the subscales provides clarity and nuance that only reviewing the full scales cannot accomplish.

5. Configural Interpretation

The final level of interpretation of the PAI relates to configural interpretation, which involves combining scales and subscales in clinically meaningful ways. Different configurations often enhance the overall meaning of the scores. Practitioners with more experience interpreting the PAI as well as extensive knowledge of psychological theory and research are likely to work quite effectively with the various combinations of scale elevations. The most basic and simplistic configuration is the 2-point code type, which evaluates two scales and/or subscales concurrently for more detailed interpretation. Each scale description in the next section includes discussion of frequently found code types.

More nuanced and potentially more meaningful configurations rationally combine and evaluate multiple scales or subscales in a theoretically driven way. For example, while the PAI does not have a scale for Narcissistic Personality Disorder (NPD), several of the scales and subscales rationally inform the theoretical and empirical conceptualization of the disorder. Evaluating the Grandiosity (MAN-G), Dominance (DOM), and Egocentricity (ANT-E) scales together can inform a potential diagnosis of NPD. This is true even though ostensibly the ANT-E subscale is related to Antisocial Personality Disorder and the Grandiosity subscale is related to mania. The experienced practitioner may have specific diagnostic or other clinically relevant questions going into the assessment that can be informed by configuring multiple scales and/or subscales in ways not presented on the profile form. Again, the development of scales and subscales to be interpreted directly as their scale names suggest allows the practitioner the freedom to reconfigure measured concepts in different, useful ways.

VALIDITY SCALES

Inconsistency (ICN)

The Inconsistency (ICN) scale is a measure of the degree to which 10 pairs of items with very similar content are rated consistently or inconsistently by the client. The 10 pairs were determined based on the items that empirically had the highest correlation during the development of the PAI. The process is not as straightforward as similar scales on personality measures that employ a true-false response format, as even items with very similar content may be slightly different when given four response alternatives, as is true on the PAI. Each of the 10 pairs of items is evaluated for absolute difference (with some of the 10 needing reverse coding for this calculation, as the expected responses on these items are opposites). The sum of these absolute differences constitutes the ICN raw score. *T* scores on ICN are interpreted in the normal, mildly elevated, and significantly elevated ranges. Profiles with normal range ICN ($T < 64$) suggest that

the client generally responded in a consistent and deliberate manner. Mild elevations ($63 < T < 73$) suggest that there was some inconsistency in responses, which could be related to several factors, including carelessness or deliberate impression management. All other scales on profiles within this range should be interpreted with caution. Significantly elevated ICN scores ($T > 72$) suggest that the measure includes true inconsistency in responses, which could be due to difficulties with reading ability, defensive refusal to follow directions, confusion, or other reasons. Profiles with ICN scores in this range should be considered invalid, and it is recommended that the practitioner not continue to interpret any other scales.

Infrequency (INF)

The Infrequency (INF) scale includes eight items that are known to be rated similarly by virtually anybody who takes the test, is paying attention to the questions, comprehends their meaning, and is responding honestly. These items were developed to be answered consistently by all respondents without being bizarre; that is, the items are extremely unlikely to be answered in a way different from expected, without being impossible or unfathomable (such as a statement that an individual's preferred writer is an extremely obscure one). Half of the items are expected to be answered *Very True (VT)*, whereas the other half are expected to be answered *Very False (VF)*. These items are distributed relatively evenly throughout the measure, in order to control for a change in response strategy throughout the test. For example, an individual who becomes fatigued during the measure and begins responding randomly would likely offer infrequently endorsed ratings on the INF items toward the end of the test. Moderate elevation on INF ($59 < T < 76$) indicate some idiosyncratic responding, and interpretations should be made cautiously. Significant elevations ($T > 75$) suggest problematic responding, and practitioners should consider not interpreting the other scales of the PAI. Morey (1996) suggested that elevations on the INF may be due to confusion, carelessness, or resistance, but it may also be due to "a tendency to answer the PAI items in a very idiosyncratic way" (p. 107). He gave an example of a respondent who interpreted a question in a way that is significantly different from most people, asserting that his favorite sporting event to watch on television is the high jump, despite having never seen the high jump on television, simply because he wanted to convey to the assessor that watching sports on television is important to him, and there was no other way to convey that on the PAI. The problem with this type of idiosyncratic thinking is that it likely colors the way this individual interprets many, if not all, of the questions, and so interpretation of the PAI must be made very cautiously, if at all.

Negative Impression (NIM, MAL, RDF)

The PAI includes several measures that can be used to evaluate those who are purposely presenting themselves in a negative manner, with more symptoms than they likely truly have. The first and most straightforward is the Negative Impression Management (NIM) scale. The NIM includes items that present likely exaggerated or negatively distorted impressions of the respondent and symptoms that are characteristically bizarre and highly unlikely to be validly endorsed in any population. Moderate elevation on

NIM ($72 < T < 84$) suggests some negative presentation, while slightly higher scores ($83 < T < 92$) often represent a deliberate distortion in the negative direction, possibly as a cry for help. Significant elevations ($T > 91$) suggest a strong possibility of either malingering or careless responding, and practitioners should consider not interpreting the PAI further. As always, this scale should be interpreted within the context of what has been clinically presented by the respondent, as elevations on this scale may correlate logically with his or her current circumstances.

The Malingering Index (MAL) was developed as a configural approach to detecting those who are more likely simulating psychopathology than legitimate clinical populations. Eight criteria are presented and determined to be present or not, with a higher number of criteria present indicating a higher likelihood of malingering. While the NIM is included on this list (two of the eight criteria relate to elevations on the NIM), other indicators are present related to elevation in Infrequency (INF) not due to Inconsistency (ICN) or carelessness (INF – ICN); elevation persecutory paranoid ideation (PAR-P) not related to hypervigilance (PAR-H) or resentment (PAR-R) feelings (PAR-P – Par-H and PAR-P – Par-R); elevations in egocentricity (ANT-E) without specific antisocial behaviors (ANT-A) present (ANT-E – ANT-A); and others. Moderate elevations ($2 < \text{raw score} < 5$) reflect the possibility of malingering, while significant elevations (raw score > 4) represent a strong likelihood of malingering. The sensitivity of this scale is stronger when the psychopathology being malingered is severe in nature (e.g., psychosis), while those faking milder forms of pathology (e.g., anxiety) may not trigger elevations on MAL.

The Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) provides another score with a similar aim to the MAL. It uses 20 different items, weighting their T scores and summing the results, to distinguish individuals who are feigning psychopathology from those who are genuinely clinical patients. The RDF was found to be effective in both community and patient samples, as those with both naive and sophisticated/knowledgeable (graduate students) attempts at malingering were consistently identified as separate from the clinical population. Even sophisticated “fakers” were unable to accurately simulate psychopathology on this complex measure, with 20 different T scores comprising its calculation. The score is a discriminant function score with a raw cutoff of 0, which equates to a T score of 59. Positive scores (raw score > 0) represent a likelihood at purposeful malingering, while negative scores (raw score < 0) represent no effort at negative self-presentation.

Positive Impression (PIM, DEF, CDF, ALC Est, DRG Est)

The PAI currently includes five separate measures of possible defensive or positive impression management, with two of these specifically related to the likelihood of underreporting alcohol and drug use. The first and most straightforward of these scales is the Positive Impression Management (PIM) scale. The PIM includes items that relate to self-favorable presentation and the denial of even minor faults. This scale can be a difficult one to interpret, as a number of populations (general, research, and even clinical populations) can answer in an honest way that could be interpreted as their overrepresenting positive psychological functioning. As a result, the PIM tends to be used more informationally than as a decision maker of whether to interpret the PAI

further or not. Moderate elevations on PIM ($56 < T < 68$) suggest that an individual is portraying him- or herself generally favorably, free of even many faults that are quite common. A score in this range could represent a purposeful strategy for denying problems (which may relate, e.g., to secondary gain related to a positive assessment outcome) or a more underlying, defensive coping strategy (which may relate, e.g., to some undercurrent of shame around anything perceived as imperfect or potentially problematic). Significant elevations ($T > 67$) suggest that the individual is denying even those common problems to which most people readily admit. Scores in this range indicate that the rest of the PAI should be interpreted within the context of understanding that the respondent is portraying him- or herself in an overly positive way and is likely underreporting problems throughout.

The Defensiveness Index (DEF) was developed as a configural approach to detecting those who are more likely presenting themselves in a positive manner than both normal and clinical populations. Similar to MAL, eight criteria are presented and determined to be present or not, with a higher number of criteria present indicating a higher likelihood of a defensive style of responding. While the PIM is included on this list, it is only included in one of the eight criteria. Other indicators include elevations in egocentricity (ANT-E), stimulus seeking (ANT-S), and obsessive-compulsive symptoms (ARD-O) with antisocial behavior (ANT-A) not involved (ANT-E – ANT-A; ANT-S – ANT-A; ARD-O – ANT-A); presence of grandiosity (MAN-G) without irritability (MAN-I), which often relates to wanting help (MAN-G – MAN-I); high treatment resistance (RXR); and others. Again, as defensiveness and positive impression management are difficult constructs to measure and interpret, given the widespread and somewhat normative employment of these strategies among different populations, the DEF is somewhat complicated to interpret. Elevations on the DEF (raw score > 5 ; $T > 69$) reflect a high likelihood that the individual is being purposefully and clearly defensive in his or her approach to the test. However, individuals who score below this range (raw score < 6 ; $T < 70$) are not necessarily being nondefensive and honest in their presentation. Sophisticated defensiveness can bypass the triggering of DEF, but significant elevation should be taken as a signal to interpret the rest of the PAI within the context that the individual is likely presenting in a defensive and overly positive manner.

The Cashel Discriminant Function (CDF; Cashel, Rogers, Sewell, & Martin-Cannici, 1995) provides another score with a similar aim to the DEF in a manner similar to the RDF. It uses six different items, weighting their T scores and summing the results, to distinguish individuals who are responding defensively from those who are being generally honest. Like the RDF, the CDF was found to be effective in both community and patient samples, effectively discriminating those with defensive styles from true mental health. Moderate elevations ($144 < \text{raw score} < 168$; $54 < T < 70$) represent some efforts toward presenting more favorably, while significant elevations (raw score > 167 ; $T > 69$) suggest that the scores on the PAI are more likely a reflection of how the individual who responded wants to portray him- or herself than a reflection of his or her true current functioning.

Because of the relative frequency of defensive responding about alcohol and drug use, even by those who respond nondefensively to the rest of the measure, some estimates of alcohol and drug use are calculated using items that relate to these behaviors

but do not directly ask about actual use. Although weighted and calculated differently, the estimated scores for alcohol use (ALC Est) and drug use (DRG Est) use the same five PAI scales—self-harming behaviors (BOR-S), antisocial behaviors (ANT-A), egocentricity (ANT-E), stimulus-seeking (ANT-S), and physical aggression (AGG-P)—all of which have been shown to be highly related to substance use. When the *T* score for the alcohol problems (ALC) and drug problems (DRG) scales are compared to the ALC Est and DRG Est scales, significant discrepancies should be noted. Specifically, when the ALC Est *T* score exceeds the ALC *T* score by 10 or more points, it is a sign of possible denial or defensiveness about alcohol use problems. When the DRG Est *T* score exceeds the DRG *T* score by 10 or more points, it is a sign of possible denial or defensiveness about drug use problems. Although no definitive conclusion can be made based on these discrepancies, they should trigger further evaluation of these specific areas, whether by reaching out to collateral informants or adding measures of subtle signs and symptoms of alcohol and drug abuse and dependence.

CLINICAL SCALES

Somatic Complaints (SOM)

The key characteristic of the SOM scale is its focus on the degree of concern about, and thus psychological reaction to, physical health matters. The subscales (discussed below) are meant to help differentiate between key somatization disorder-type presentations. Even mild elevations on the SOM scale can indicate that physical wellness is a central concern for the individual and his or her overall functioning. The SOM scale includes the perception of how impaired or compromised the individual is because of physical or health concerns. Moderate elevation on SOM ($59 < T < 70$) suggests some bodily or health concerns, and is relatively common in older adults and those with known medical illnesses, for whom bodily concerns should necessarily be present. Significant elevations ($T > 69$) represent individuals who feel they are in poor health and have complex medical problems that are difficult to understand, explain, and treat. When SOM is extremely elevated ($T > 86$), the individual is likely preoccupied with his or her health concerns, and the concerns are likely to overrun much of his or her life. Such individuals will complain of fatigue and inability to function because of their medical problems, and their self-image will be largely influenced by taking on a physically ill and/or patient role in the world.

The Conversion (SOM-C) subscale focuses on unusual sensory, perceptual, and motor problems often found in conversion disorder. Included in these problems are numbness, paralysis, and dramatic problems with hearing and vision. It is important to note that while symptoms like those on this scale are rare in the general population, an elevation on this scale may not necessarily indicate conversion disorder, and should not be used in isolation of other clinical information to diagnose it, as there are some medical (especially neurological) illnesses that present with these symptoms. Elevations on SOM-C ($T > 69$) indicate that the individual is presenting with at least one of these neurological, pseudoneurological, or musculoskeletal symptoms and that he or she is concerned about its impact on functioning. Extreme elevations ($T > 94$) indicate that there is a possibility of somatic delusion, and this symptomatic hypothesis should be evaluated.

The Somatization (SOM-S) subscale focuses on more vague, general, and diffuse physical complaints than SOM-C, such as headaches, fatigue, pain, and gastrointestinal problems. For those with elevation on SOM-S ($T > 69$), these broad, general physical problems get in the way of everyday functioning, enough that they are concerning to them. They are dissatisfied with their own health and well-being, as well as with the fact that their many minor physical problems are difficult to treat (or have not yet appropriately been treated). They often revel in discussing their physical complaints with others, to an exhausting degree.

The Health Concerns (SOM-H) subscale focuses on an individual's degree of focus and preoccupation on his or her own health. It is not a measure of the degree to which an individual feels impaired as much as it is a measure of how much mental energy is used on thinking about the health problems. Individuals with elevations on SOM-H ($T > 69$) do indeed feel that they are less healthy than others around them, but they also feel that their health problems are complex and difficult to treat, and they spend much energy trying to treat them. These individuals are preoccupied with their own health, and much of their identity may be wrapped up in this preoccupation. While genuine health concerns can elevate SOM-H, the hypothesis of hypochondriasis should be considered and explored.

Frequent Code Types

Scales that are likely to be elevated along with SOM are Anxiety (ANX), Depression (DEP), and Alcohol Problems (ALC). When ANX and DEP are elevated along with SOM, the pattern relates to individuals who are currently in acute distress, particularly around their physical functioning. While their physical problems are significantly impacting their lives negatively, how the disruption manifests itself depends on the other scale, either as tense and nervous (ANX) or as unhappy and listless (DEP). Commonly, all three (SOM, ANX, and DEP) are elevated in individuals who are currently experiencing significant diffuse distress, with some physical focus. When SOM and ALC are both elevated, it is likely that some of the physical complaints are related somehow to the history of alcohol use, including residual effects of heavy use or current effects of alcohol abuse. A vicious cycle, this pattern can reinforce itself, as individuals may be using alcohol to cope with their physical distress. Additionally, some individuals are displaying and coping with emotional distress in any way possible not to overtly feel negative feelings; as such, their distress may be focused on physical problems, and they may be using alcohol to cope with any underlying issues as well.

Treatment Implications

Elevation on SOM reflects two major possibilities: the presence of real physical and medical problems that are causing anxiety and serve as a major focus of attention and the possibility that physical concerns are the way the individual experiences distress. The former possibility—that there is a preoccupation with physical functioning related to real medical problems—would reflect a need for the treatment to focus on minimizing the catastrophization and making sure the preoccupation and likely resultant anxiety about the physical concerns are not in fact exacerbating the medical problem. The latter possibility—that the individual experiences distress in physical manifestations rather than psychological or emotional ones—presents a significant treatment challenge. Experiencing distress physically often comes with difficulty

understanding that psychological factors may play a role in problems as well as a general lack of emotional understanding and psychological mindedness. Mental health professionals are often a last resort for these individuals, after they have been evaluated by many different medical professionals, and a therapeutic alliance can often be difficult to forge because of the suspiciousness and even cynicism with which individuals with SOM elevations may dismiss therapists, expecting to hear that all of their physical complaints are somehow psychologically fabricated. Individuals with elevations on SOM may benefit initially, while building rapport, from less insight- or even psychologically oriented treatments, such as being taught relaxation techniques.

Anxiety (ANX)

The ANX scale focuses on the worry, tension, and general negative affect related to anxiety in general. It does not focus on specific behaviors related to anxiety disorders, which are addressed on the ARD (Anxiety-Related Disorders) scale, but rather the general presentation of anxious and negative symptoms that accompany many different mental health concerns. That is, elevations on ANX may indeed represent the presence of an anxiety disorder, but alternatively such elevations may be related to a diagnosis of depression, in which anxiety can play a central role. Moderate elevations on ANX ($59 < T < 70$) indicate the presence of stress, which may be situationally related. At this level, the clinician should investigate the individual's current circumstances, to see if the stress and tension presented is a realistic and normal reaction to stressors. Significant elevation ($T > 69$) suggests significant anxiety and tension that likely gets in the way of normal, everyday functioning in some way. Individuals with this degree of ANX elevation are worried and tense much of the time, and elevation in this range means that at least one of the subscales is likely elevated, so they should be examined for more specific information about their felt experience. At extreme elevation of ANX ($T > 89$), anxiety is likely impairing the individual's functioning, as worry, rumination, and tension eclipse the ability to carry out many meaningful tasks. Individuals at this level may suffer from Generalized Anxiety Disorder.

The Cognitive Anxiety (ANX-C) subscale focuses on the cognitive components of worrying and ruminating about potential harm, negative events, and danger. Individuals with elevations on ANX-C ($T > 69$) are overly concerned about current circumstances, often over which they have little or no control. Their worry takes the form of thinking things over and over, such as anticipating problems or being on the lookout for adverse situations, to the point that it impairs their attention. Their worried thoughts tend to overwhelm them and take precedence over logical or rational thoughts.

The Affective Anxiety (ANX-A) subscale focuses on the general feelings of tension, nervousness, and fear. These feelings tend to be somewhat persistent and dispositional, more than situationally related. Individuals with elevations on ANX-A ($T > 69$) are dealing with a great deal of stress and worry and have difficulty calming down and relaxing. In the absence of elevation on the other two ANX subscales, elevation on ANX-A is most likely related to the free-floating, long-standing dispositional anxiety related with generalized anxiety disorder. As with all the scales and subscales, it is important to interpret ANX-A within the context of what is going on in the individual's

life and environment, as heightened feelings of anxiety may be a normal, expected reaction to current circumstances.

The Physiological Anxiety (ANX-P) subscale focuses on the physical and physiological manifestations of anxiety, including quick breathing, racing heartbeat, sweaty palms, and feeling dizzy. Although the individual may not relate these symptoms to anxiety, especially if ANX-C and ANX-A are low, the symptoms tend to cluster around the physiological expression of anxiety. Individuals with elevation on ANX-P ($T > 69$) are experiencing not only the physical symptoms of anxiety, such as those just listed, but they are likely to be experiencing other somatic or physical symptoms as well. When this subscale is elevated, the SOM scale and subscales should be evaluated together with it.

Frequent Code Types

Scales that are often elevated along with ANX include Anxiety-Related Disorders (ARD), Somatic Complaints (see “Somatic Complaints [SOM]” section above), Depression (DEP), Alcohol Problems (ALC), and Drug Problems (DRG). Anxiety as a symptom cuts across many different psychopathological presentations, and the broad anxiety tapped by the ANX lends to it being elevated along with other scales. When elevated along with ARD, there is a strong possibility the individual will meet criteria for a specific anxiety disorder. A review of the ARD subscales can help identify the quality of the anxiety, and the ANX scale helps underscore just how intrusive the anxiety is in the individual’s life and functioning. When ANX and DEP are elevated concurrently, the individual is noticeably unhappy and moody and is specifically aware of the need to improve his or her mental and emotional state. As anxiety and depression often occur comorbidly, and even more so the symptoms related to these syndromes, it is not unusual to see the diffuse anxiety related to ANX and the overall sadness, moodiness, low self-esteem, and pessimism of DEP present in the same individual. When ANX is elevated along with ALC or DRG, a number of different mechanisms may be at play, and often more than one is occurring simultaneously. Alcohol and drug abuse may serve the purpose of reducing the tension related to the anxiety. Alternatively, an individual may be tense, guilty, and anxious about the alcohol or drug abuse. The alcohol or drugs may actually be serving to heighten the worry and nervousness related to an already anxious individual, and both likely cause difficulties interpersonally or occupationally, which then serve to reinforce the problem (heightening anxiety and increasing alcohol or drug abuse). The relationship between anxiety and substance misuse is complex and nuanced, and when ANX is elevated along with ALC or DRG, further assessment of this pattern is warranted.

Treatment Implications

As related to significant immediate felt distress, individuals high on ANX are likely quite motivated to undertake therapeutic treatment to improve their functioning. Diffuse and broad feelings of anxiety are uncomfortable enough that most people with elevations on ANX will openly accept help. However, the subscales can help tailor treatment to what will likely be most effective. For example, individuals with elevated

ANX-C are good candidates for cognitive therapy, in order to challenge, alter, and restructure their anxious cognitions. Those with elevation on ANX-P may be more receptive, at least in the immediate term, to psychophysiological interventions, such as deep breathing and biofeedback. These individuals, along with those elevated on ANX-A, may also be good candidates for medication treatment.

Anxiety-Related Disorders (ARD)

The ARD scale differentiates the behavioral symptoms of three different specific anxiety disorders, in contrast to the ANX scale, which focuses on the broad, general experience of anxiety. While elevations on ARD can be interpreted as behavioral manifestations of anxiety disorders, it is less important to interpret this scale than to examine the three subscales that comprise it. The three subscales represent distinct, disparate anxiety disorders, and as such the ARD itself is not as useful (in contrast to the rest of the clinical scales, which represent overarching factors made up of components represented by the subscales). While Morey (2007) does present interpretive criteria for this scale, it is recommended that interpretation of ARD fall directly to its subscales.

The Obsessive-Compulsive (ARD-O) subscale focuses on both the behavioral and thought features of obsessive-compulsive disorder, such as ruminative fears and ritualistic behavior, and the cognitive styles and behaviors related more to obsessive-compulsive personality disorder, such as perfectionistic tendencies and extreme detail orientation. Both of these clusters of symptoms are represented on this subscale, such that those with specific obsessions or compulsions are equally able to elevate the scale as those with rigidity, constriction of emotion, and restrictive, rule-bound beliefs. Elevations on ARD-O ($T > 54$) represent efforts by the individual to impose order and control in response to anxiety. The degree to which these efforts are prevalent in an individual's functioning is represented by the degree to which the scale is elevated, with moderate elevation ($54 < T < 66$) in clinical settings suggesting some rigidity and rumination, significant elevation ($64 < T < 75$) suggesting rigidity and uncompromising personal conduct that likely impairs normal functioning, and extreme elevation ($T > 74$) suggesting clear rigidity and rumination that manifests as disruptive obsessive thinking or compulsive behavior. The higher the T score on this scale, the more likely that rigidity and ruminative thoughts are intruding on the individual and his or her everyday functioning.

The Phobias (ARD-P) subscale focuses on some of the most common situations and contexts that tend to produce phobic reactions in those with specific phobias. Included in these are heights, social situations, public transportation, confined spaces, and specific objects. Individuals with significant elevations on ARD-P ($T > 69$) not only have specific fears associated with one or more of the situations or objects just listed, but they tend to defensively avoid them and be hypervigilantly aware of their potential presence in their environment. High scores on ARD-P suggest that the phobias are interfering with their everyday living and functioning. An interesting note on the ARD-P subscale is its interpretive value at depressed scores. Significantly low scores on ARD-P ($T < 36$) suggest that an individual does not exhibit fear even when appropriate, which can lead to recklessness or even dangerous behavior.

The Traumatic Stress (ARD-T) subscale focuses on the experience of a traumatic incident or stressor that fundamentally changed the individual in some way, with lasting, current effects. The scale is not intended to be a direct assessment of posttraumatic stress disorder, which will likely elevate this and other scales, but it evaluates circumstances that are common in many clinical populations. Moderate elevation ($64 < T < 76$) represents the presence of a significant traumatic event in the individual's past, one that changed him or her and continues to be a source of concern in some way. Individuals with elevations are reporting symptoms like nightmares and sudden, intense anxiety, in addition to constant thinking about some past traumatic event. Significant elevations ($T > 75$) and especially marked elevations ($T > 89$) signify that the traumatic event, as well as its lasting effects, is the primary focus of concern in the individual's life. At these elevated levels, the likelihood of posttraumatic stress disorder increases, and because the PAI does not do so, the nature of the specific traumatic event should be explored further by the clinician.

Frequent Code Types

Although most often elevated with Anxiety (see the "Anxiety [ANX]" section above), ARD is also commonly elevated along with Depression (DEP) and Borderline Features (BOR). When elevated with DEP, this pattern suggests an individual who has significant stress and recognizes that it is problematic, but low levels of energy, hopelessness, and low self-esteem can impede the ability for the individual to reach out for and accept help. They may be difficult to engage in treatment, even though they are in significant subjective distress. The pattern of elevation of both ARD and BOR suggests that stressors, historical and current, have supported an interpersonal and emotional style that is tormented, chaotic, and untrustworthy. Anger and resentment can replace neediness quite quickly, and ARD-T is often elevated, as traumatic stress has helped the individual develop and sustain the problematic, distress-inducing interpersonal style. These individuals are tense and hypervigilant, constantly on the lookout for being disappointed, mistreated, or abused. They are constantly nervous about being abandoned or let down.

Treatment Implications

Because of the nature of ARD, the implications for treatment are more significantly related to the subscales than the overall scale. The overall scale may be an indication, similar to ANX, of likely motivation for treatment, as higher elevations are likely more and more uncomfortable to the individual being evaluated. Specific treatments, though, will be related to the degree to which the individual exhibits symptoms consistent with each of the three disorders represented by the subscales. For example, obsessive-compulsive disorder (ARD-O) often responds to more structured, cognitive and behavioral treatments. Specific phobias (ARD-P) often respond to behavioral intervention like flooding, or trauma-related disorders (ARD-T). Traumatic stress (ARD-T) can lead to different outcomes, including acute stress disorder, posttraumatic stress disorder, and adjustment disorder, if the traumas are specific events like natural disasters or violent crimes, but they can also lead to interpersonal difficulties and even personality pathology, such as borderline personality disorder, if the traumas are

related more to significant relationships. These different outcomes would require different types of treatment, and personality pathology may lead to significant resistance within the therapeutic relationship, as interpersonal trust is often compromised after interpersonal trauma.

Depression (DEP)

The DEP scale focuses on the major components of depressive syndromes, organized in its subscales by cognitive, affective, and physiological signs of depression. Included are helplessness and pessimism, sad and distressed feelings, and decreased energy and motivation, among other common symptoms. Moderate elevations on DEP ($59 < T < 70$) indicate the presence of some depressive symptomatology, such as self-doubt and pessimism, at least on occasion. At this level, the clinician should investigate the individual's current circumstances, to see if there are current logical and realistic reasons for the individual to be experiencing such negative feelings; they may be natural, realistic reactions to a legitimately negative situation. Significant elevation ($T > 69$) suggests significant sadness, dysphoria, and negative cognitive attributions, such as explaining positive outcomes in terms of luck or circumstance and negative outcomes with enduring personal failings. At extreme elevations of DEP ($T > 79$ and especially $T > 94$), there is an increased likelihood that the individual is currently experiencing a major depressive episode. Such individuals feel helpless, hopeless, and worthless; their dysphoria is pervasive and inhibits normal functioning; and they display vegetative symptoms of depression, including changes in sleep and appetite, low libido, and psychomotor slowing.

The Cognitive Depression (DEP-C) subscale focuses on the cognitive components of thinking one is worthless, helpless, and generally inadequate. Individuals with elevations on DEP-C ($T > 69$) think of themselves as unable to adequately meet the demands of life and helpless in the face of stressors. They attribute success to luck, external forces, or temporary factors, whereas they attribute failure to pervasive or permanent personal inadequacy and failings. DEP-C is highly linked to self-esteem (how a person feels about him- or herself in general) and self-efficacy (how competent a person feels he or she is to effect change in his or her life). High DEP-C is related to general pessimism in life as well.

The Affective Depression (DEP-A) subscale focuses on the general feelings of distress, dysphoria, and sadness. The subscale is highly inversely related to satisfaction with life and as such is a good indicator of an individual's desire to change his or her current circumstances or coping strategies. Individuals with elevations on DEP-A ($T > 69$) are currently distressed, gloomy, and sad, and exhibit anhedonia (a loss of pleasure in things they once enjoyed). They have lost interest in activities that previously interested them, and higher scores on DEP-A reflect the degree to which their general dysphoria is interfering with their ability to function in everyday life. As with all the scales and subscales, it is important to interpret DEP-A within the context of the individual's current circumstances, as there may be logical reasons the individual is feeling down and depressed.

The Physiological Depression (DEP-P) subscale focuses on the vegetative signs and physical manifestations of depression, including changes in appetite or weight, altered sleep pattern, and decreased motivation, interest, and energy level in general. Individuals with elevation on DEP-P ($T > 69$) report the vegetative signs above, as well as psychomotor slowing, decreased libido, and general malaise. Morey (2007) explained that the symptoms measured by DEP-P are ones most often targeted in studies of antidepressant medication; as such, the DEP-P may inform recommendations for medication treatment of depression.

Frequent Code Types

Depression, and especially some of its related signs and symptoms, is frequently found comorbidly with symptoms of other psychological problems. Among the most common patterns of elevation along with DEP are anxiety (see the “Anxiety [ANX]” section), anxiety-related disorders (see “Anxiety-Related Disorders [ARD]” section), schizophrenia (SCZ), borderline features (BOR), and alcohol and drug problems (ALC and DRG, respectively). Elevations in both DEP and SCZ signify that an individual has marked difficulties with his or her thinking and thought processes, accompanied by sadness and general distress. It is likely that the problems in thinking are negatively affecting interpersonal and occupational situations, as well as attention and concentration more generally, and these disruptions are causing distress. The combination of helplessness, hopelessness, and potentially poor judgment associated with thought problems place these individuals at higher risk for harming themselves. When DEP and BOR are elevated together, it marks an individual who, while often moody and labile, is currently disappointed, disillusioned, and feeling rejected or abandoned by others. It is likely that repeated interpersonal problems have left the individual pessimistic and hopeless about the world and other people in it, and thoughts are likely centered on how everybody will eventually abandon or let him or her down. While in acute distress, such individuals can be resistant to believing that a therapist can be helpful, as a therapist is another attachment figure who will inevitably fail them.

When DEP is elevated along with ALC or DRG, the individual is unhappy, pessimistic, and likely somewhat desperate. The relationship between depression and substance use is complex and nuanced, with directionality often hard to discern. Depression can lead to substance use as an escape from the pain and dysphoria that can encompass a depressed person’s thinking and feeling. Alternatively, substance abuse can lead to social and interpersonal failings and difficulties, which can certainly reinforce depression. Additionally, underlying factors, such as early trauma or a tumultuous early home life, may influence both depression and substance abuse concurrently. While directionality may have been present at the beginning of the problems, often it becomes less salient and important as the problems develop and begin to reinforce each other in a feedback loop. With elevations on both DEP and either ALC or DRG, individuals can be difficult to engage. If the alcohol or drug use is current, they may not be ready to engage in helping behaviors, despite knowing that they are in distress and have problems. These individuals are likely not hopeful about

improvement. This configuration is also commonly associated with elevations in BOR and Suicidal Ideation (SUI), which should be examined closely in these individuals.

Treatment Implications

Depression is one of the most widely studied psychopathologies and has been found to be generally responsive to a number of different types of therapeutic treatment. As a general indicator of subjectively felt distress, the DEP is a good indicator of desire to change and thus motivation for treatment, unless it is so elevated that motivation disappears and apathy takes over as a major mode of being. Like ANX, the subscales can help determine what kinds of treatment may be useful, with elevation in DEP-C suggesting that the individual would likely benefit from cognitive techniques and elevation in DEP-P suggesting that psychopharmacological intervention may be useful. In general, though, a therapeutic relationship with specific interventions will likely benefit any individual who is elevated on DEP.

Mania (MAN)

The MAN scale was originally designed to focus on the hallmark features of a manic episode, but it also includes components of hypomania. Some of these components are elevation in mood, expansiveness, grandiosity, heightened activity level, and irritability. Because of a low base rate of manic and hypomanic symptoms even in clinical settings, the threshold for elevation interpretation is somewhat lower than with most of the other PAI scales. Moderate elevations on DEP ($54 < T < 65$) represent individuals who are highly active and outgoing, generally confident and bold, and at times impatient, assertive, and potentially aggressive. Significant elevation ($T > 64$) relates to an increase in irritability, grandiosity, and volatility of mood. Individuals with significant elevation on MAN are restless and impulsive, often not thinking through decisions and acting rashly. At extreme elevation of MAN ($T > 74$), there is an increased likelihood that the individual is currently experiencing a manic or hypomanic episode. Such individuals are impulsive to a destructive degree, with little ability to delay gratification, grandiose possibly to a delusional level (which should be examined further), and overly exuberant in their activity level. At this marked elevation of MAN, individuals' interpersonal functioning is often compromised by their grandiosity, irritability, and racing thought process.

The Activity Level (MAN-A) subscale focuses on the increase in speed and quantity of both thoughts (including flights of ideas) and behaviors (including psychomotor agitation). Individuals with elevations on MAN-A ($T > 64$) are noticeably accelerated in their thinking and behavior, participating in many activities but doing so in a disorganized, confused manner. Their high energy level makes it difficult for them to focus and complete tasks, often impairing their ability to sustain attention and persist. Marked elevations on MAN-A ($T > 74$) often signify an activity level that makes individuals both confused and confusing to others. Their line of thinking is difficult to follow, and the likelihood of their currently being in a manic state is high. Low scores on MAN-A ($T < 30$) reflect lethargy, apathy, and indifference, which are often found in individuals currently experiencing a depressive episode.

The Grandiosity (MAN-G) subscale focuses on a heightened self-evaluation, ranging from how well a person thinks he or she performs everyday tasks to the belief that the person has special and unique talents and skills. The subscale is highly related to self-esteem, and low scores can signify poor self-evaluation and low self-esteem and self-efficacy, even in those individuals who are not currently depressed. Individuals with moderate elevations on MAN-G ($59 < T < 71$) are optimistic and self-confident, while those with significant elevations ($T > 70$) believe they excel at nearly everything they do. At these levels, individuals are often seen as grandiose and narcissistic, believing they have special talents and skills that should be appreciated by others. It is important to interpret these elevations in the context of the individual's clinical presentation, as the person may have found a way to channel his or her grandiosity into productive avenues. For example, an executive who has worked his or her way up in a company may have found good use for inflated confidence. As with most symptoms, the questions of whether and how the trait is impairing the individual's functioning is key to understanding its relative importance in a potential clinical picture.

The Irritability (MAN-I) subscale focuses on the volatile mood component of mania. This subscale assesses impatience, ability to tolerate frustration, and rapid shifts in mood. Individuals with moderate elevation on MAN-I ($59 < T < 71$) are impatient and demanding, blaming others who do not cooperate with them. Significant elevations ($T > 70$) suggest extreme volatility, especially in the face of frustration, disagreement, or perceived noncooperation by others. Their mood can swing abruptly, and their judgment in situations of frustration tends to be quite impaired. At this level of elevation, an individual's interpersonal relationships will likely suffer.

Frequent Code Types

The MAN scale is not often elevated along with other scales. Although individuals with manic and hypomanic episodes are most likely to have depressive episodes as well, because of the nature of the PAI, it would be unlikely to see elevations on MAN and DEP *at the same time*, as respondents are unlikely to have the features of both concurrently. MAN can be elevated concurrently with the Borderline Tendencies (BOR) scale. When this pattern emerges, the individual is both irritable and impulsive. Mood swings are prominent, and the heightened energy reflected in the MAN scale renders the individual even more reactive to minor stimuli. That is, very small (or even innocuous) events can trigger extreme emotional reactions from these individuals. They are unpredictable and tend to act out, especially against those close to them.

Treatment Implications

Elevations on the MAN scale are generally not hopeful indicators for psychotherapeutic treatment. Symptoms tapped on this scale are not generally too distressing to the individual being evaluated, though they may be markedly distressing to others around them. Different configurations of the subscales can also indicate different things. For example, if the subscales indicate a current manic or hypomanic episode, medication treatment and compliance with it should be a major focus of treatment. Grandiosity (MAN-G), especially in its extreme elevation, signifying narcissism, may indicate the presence of fragile self-esteem and outward defense against it. Treatment of such

individuals can be complicated, and a therapeutic alliance that focuses on care and delicateness around the client's self-esteem is generally necessary. When MAN-A is elevated, the individual will likely have very limited patience for slow-paced, nondirective treatment.

Paranoia (PAR)

The PAR scale was developed in response to the inherent difficulty in evaluating paranoid ideation in a self-report measure, as those with paranoid symptoms are necessarily suspicious and thus defensive about responding. As such, in addition to including some overt paranoid symptoms, like suspiciousness, jealousy, and the feeling that others are expending effort to undermine them, the PAR also includes items related to the experience and mode of interacting with the world of those with paranoid ideation. This experience includes their awareness and alertness to potential threats in their environment, sensitivity to any perceived slights, attribution of any negative circumstances to the belief that they are being treated unfairly or are being targeted and victimized, and the tendency to be resentful and hold a grudge against others they view as having victimized or targeted them. Moderate elevations on PAR ($59 < T < 70$) represent individuals who are wary and tentative in relationships, skeptical and at times cynical about the motivations of others. Significant elevation ($T > 69$) relates to suspiciousness and hostility, expending energy to monitor their surroundings for potential threats and abuses. These individuals are jealous of others who they feel are treated better than they are and resentful toward those who they think are being unfair toward them, and they do not relinquish grudges easily. At extreme elevation of PAR ($T > 83$), there is an increased likelihood that the individual is experiencing specific paranoid delusions, including persecution or grandiosity. At this level, the specific content of potential delusions should be deliberately evaluated.

The Hypervigilance (PAR-H) subscale focuses on the guardedness and suspicion with which individuals approach others that they do not know well. The subscale taps constructs ranging from skepticism toward others to preoccupation with the potential for threat, spending a great deal of time and energy scanning the environment for potential abuses, exploitation, and other dangerous interaction. Individuals with elevations on PAR-H ($T > 59$) are reluctant to get close to others, always aware of the potential that others will manipulate or abuse them. Marked elevations on PAR-H ($T > 70$) relate to a preoccupation with the potential threat of others, mistrusting the motivation and questioning the potential ulterior motives of others. These individuals are constantly scanning the environment for threat, and they are extremely sensitive to even small indicators of interpersonal exploitation. They are quick to feel insulted by others and to attribute any interpersonal conflict to the malicious intents of the other person. Higher scores on PAR-H increase the likelihood of paranoid personality disorder. Low scores on PAR-H ($T < 40$) reflect individuals who are extremely trusting in relationships, potentially to the point of being vulnerable to exploitation and manipulation.

The Persecution (PAR-P) subscale focuses on the degree to which individuals believe that others are actively trying to undermine them and that they are being treated inequitably. Individuals with elevations on PAR-P ($T > 69$) are quick to believe

others are making efforts to get in their way, sabotage their efforts, or block them from succeeding. These efforts of others are concerted and purposeful, and they result in failures, lack of opportunities, and general problems in work and life. At marked elevation ($T > 85$), clinicians should investigate the possibility of paranoid delusions, as the beliefs that others are undermining them are so striking that they likely represent concerns significantly based outside of reality.

The Resentment (PAR-R) subscale focuses on the bitterness, resentment, and negativity harbored by individuals who feel they have been treated unfairly in life or have been wronged by others. This subscale focuses on the bitter feelings rather than the way these feelings are likely expressed. Individuals with moderate elevation on PAR-R ($59 < T < 71$) are sensitive to others and are easily hurt or insulted by how others talk to them, behave, or even succeed. Significant elevations ($T > 70$) suggest that the individual feels that his or her own failures are attributable to others treating him or her unfairly, favoritism, or neglect, and he or she tends to hold onto grudges, whether fair or not. These individuals tend to credit luck or favoritism for the successes of others. As scores on PAR-R increase, individuals are more and more preoccupied with these feelings of resentment and bitterness. Examining other scales, such as Dominance (DOM) and Aggression (AGG), can inform the clinician how the individual is likely to express his or her resentment, such as in overtly aggressive (verbal or physical) or passive-aggressive ways.

Frequent Code Types

The most common code type pattern seen with elevated PAR is elevation in Schizophrenia (SCZ). This pattern aligns closely with the presentation of paranoid schizophrenia, with prominent confusion of thought and specifically hostile, suspicious delusions. These individuals are likely to have many life complications secondary to their difficulties in thinking, most notably in the interpersonal domain. Making and maintaining relationships with others is extremely difficult, both because of their own suspiciousness and resentment and also because of others having difficulty following their train of thought.

Treatment Implications

Elevations in PAR are often a sign that a therapeutic alliance will be extremely difficult to form, and even if it is formed, it will be difficult to maintain. If PAR is elevated and reflecting true paranoid delusions, then medication treatment is generally warranted. In therapeutic treatment, though, a major focus, whether the individual's thinking is truly delusional or not, needs to be on building trust. By its very nature, elevation in PAR comes with suspiciousness and a lack of trust in others, which makes building a relationship difficult.

Schizophrenia (SCZ)

The SCZ scale was developed to evaluate several disparate facets of schizophrenia. As such, similar to the ARD scale, because the subscales measure such different constructs, interpretation of the full SCZ scale may not be very useful. That is, individuals

with very different presentations (which may manifest as elevations different subscales) may score similarly on SCZ, and a single interpretation of this score will not characterize all of them accurately. Thus, it is recommended that interpretation be focused on the subscale level rather than the overall SCZ. However, in general, significant elevation ($T > 69$) most often relates to at least some bizarreness of thinking or behavior, such as odd thinking, alienation, aloofness, or confusion, all of which can impair attention and concentration. More specific information should be gleaned from the subscales, though.

The Psychotic Experiences (SCZ-P) subscale focuses on the specific positive symptoms of schizophrenia, including unusual sensory or perceptual experiences (hallucinations) and unusual beliefs that are held despite evidence that they are inaccurate (delusions). The subscale focuses on these very clear and distinct markers of psychotic disorders. Individuals with moderate elevations on SCZ-P ($59 < T < 71$) have unusual or eccentric ideas that likely strike others as odd. Marked elevations on SCZ-P ($T > 70$) relate to bizarre sensory and perceptual experiences and beliefs that may be delusional. These individuals are actively reporting the positive symptoms of psychotic disorders, and as scores get higher, there is a higher likelihood that individuals are actively experiencing a psychotic episode.

The Social Detachment (SCZ-S) subscale focuses on the common negative symptoms of schizophrenia, including social isolation, lack of interest in others, and flattening of emotional responsiveness. Individuals with elevations on SCZ-S ($T > 69$) tend to isolate themselves because of discomfort with closeness and contact with others, lack of real interest in relationships, and uneasiness with emotions, which are often elicited in interpersonal contexts. These individuals tend not to experience deep or strong emotions and have difficulty understanding the emotions of others, which further makes them uncomfortable with social interaction. Further, they often do not understand the nuances and subtle cues that accompany social interaction. The result is most often extreme isolation.

The Thought Disorder (SCZ-T) subscale focuses on the confused and unusual thought processes (as opposed to the unusual thought content measured by the SCZ-P) that can in turn render attention, concentration, and communication with others quite difficult. Mild elevations on this subscale most often reflect difficulties with concentration and making decisions, difficulties that are common in many psychopathological presentations. Thus, mild elevations may reflect less psychotic thought process and more cognitive outcomes of other disorders. Individuals with elevation on SCZ-T ($T > 69$) have thought processes that are marked by confusion and loose associations, which makes communicating with them somewhat to quite difficult. Their confused thinking is often manifested in self-expression that is difficult to follow, and it can also lead to behavior that others would find unusual and bizarre.

Frequent Code Types

In addition to Paranoia (see the “Paranoia [PAR]” section above), the SCZ scale has somewhat commonly been found to be elevated along with Borderline Features (BOR) and with Alcohol and Drug Problems (ALC and DRG, respectively). The pattern of elevated SCZ and BOR suggests profound confusion in the individual in terms of

thinking, feeling, and relationships in general. Although directionality is difficult to evaluate for these individuals, anger, bitterness, and resentment about interpersonal relations may make thinking unclear, just as problems with clear thinking can impair interpersonal relations in a way that breeds resentment and bitterness. As always, underlying complex dynamics may be similarly responsible for both sets of symptoms. However the symptoms are related, these individuals are certainly currently unhappy and in distress, and they are prone to impulsivity and poor judgment.

When elevations in SCZ are paired with elevations in ALC or DRG, individuals are notably confused in their thinking, with a history of or current substance abuse, and tend to be socially isolated. Similar difficulties arise with the evaluation of directionality, as substance abuse can cause impairment in thinking, and impairment in thinking aligns with poor judgment that can lead to substance abuse; notably both illogical thinking and problems related to alcohol or drug abuse can lead to and stem from difficulties in interpersonal domains. The judgment of these individuals is most often impaired, and their behavior can be erratic and confusing, though often in the service of self-isolating.

Treatment Considerations

Elevations on SCZ pose some obvious and some not-so-obvious concerns for treatment. When elevation is primarily related to SCZ-P and/or SCZ-T, psychopharmacological intervention is generally a clear option for treatment, along with psychosocial interventions aimed at medication compliance. The confusion and positive psychotic symptoms associated with elevations on these subscales, though, can make organizing and motivating individuals enough to comply with treatment difficult. An additional, notable concern is related to elevations on SCZ-S, which are associated with clients feeling guarded and strikingly uncomfortable within the interpersonal domain of treatment (among the discomfort with many different interpersonal domains). Clinicians must balance clarity, directness, and predictability, all of which will help these individuals feel more comfortable, with an environment of support and not-too-ambitious interventions. That is, interventions aimed at improving functioning of individuals with elevation on SCZ, and especially on SCZ-S, must be small in magnitude, repeated often, and applied steadily throughout treatment. These individuals can view interventions that are too ambitious as pushy or scary.

Borderline Features (BOR)

The BOR scale is similar to the SCZ scale in that it was developed to measure distinct characteristics of a specific diagnosis that is complex and multifaceted, and that has a specific definition in the field and literature. As such, examination of the subscales is critical in understanding the meaning of the BOR scale. Moderate elevations on BOR ($59 < T < 70$) often represent individuals who are somewhat moody, inconsistent, and have some difficulty in relationships. Significant elevation ($T > 69$) relates to extreme and rapid swings of mood, with features of anxiety, depression, anger, and irritability. These individuals are impulsive and feel easily hurt or betrayed by others, and when they do, they have a tendency to lash out. An extreme elevation of BOR

($T > 90$) indicates an increased likelihood that the individual meets criteria for borderline personality disorder, having significant interpersonal problems, feeling angry, resentful, and overtly anxious or depressed, and enacting self-destructive behaviors. Again, it is important to put interpretive primacy on the subscales for BOR, as different constellations of reported problems can elevate the overall scale.

The Affective Instability (BOR-A) subscale focuses on the hallmark emotional lability of borderline personality disorder, the tendency to experience rapid onset of intense, usually negative emotions, unpredictably. The emotions involved in this lability are most often anger, anxiety, depression, and irritability. Rather than tapping the specific content of the moods, the BOR-A is a measure of the rapid onset and shifts among moods. Individuals with elevations on BOR-A ($T > 69$) tend to experience sudden, uncomfortable shifts in their emotions, easily and rapidly changing to anger or some other negative feeling. The shifts will seem random and sudden to others, and these individuals may seem to have anger management problems or extreme irritability and hypersensitivity. These individuals, especially at higher scores, are moody and difficult to get along with.

The Identity Problems (BOR-I) subscale focuses on two characteristics related to the inability to define oneself and hold onto that sense of identity, the sense of who one truly is. The first characteristic is the level of directed purpose and clear goals an individual has for him- or herself. This sense of knowing what the individual wants and how he or she is going to achieve it is key to understanding who he or she is. The second characteristic related to unstable identity is the profound need to have others close, helping define the individual and who he or she is, and fearing that he or she will be abandoned by those close to him or her. Individuals with elevations on BOR-I ($T > 69$) are uncertain about who they are and what they want in life. While this is expected during the identity development phase of adolescence, adults at this level are significantly more uncertain about who they are and what they want than they should be at an adult age. These individuals rely on others to help them make decisions for themselves and define who they are, and when BOR-I is markedly elevated ($T > 80$), they are likely to rapidly and repeatedly shift their self-definition, what they want out of life, and who they want to be. These individuals do not know what they want in life and how to get it, and even though in discrete moments they may explain passionately what they want, they will change these ideas without clear reason or impetus.

The Negative Relationships (BOR-N) subscale focuses on the tendency for individuals with borderline personality disorder to become involved in relationships that are characterized as “stormy,” with intensity and drama involved. Negative relationships tend to be with friends, family, partners, and coworkers, and the same dynamics can certainly play out with a therapist. Individuals with moderate elevation on BOR-N ($69 < T < 81$) are involved in stormy relationships and likely have a history of problems in attachment with others. They tend to become close to others quickly and easily, but due to their extremely high expectations of others, they are soon let down and disappointed by them. Significant elevations ($T > 80$) reveal bitterness and resentment about having been exploited and disappointed by others in past relationships. They feel easily offended, and they are extremely sensitive to even small slights from others, often attributing these small offenses to larger patterns of betrayal. They also tend

to attribute negative motivation to even neutral stimuli, such as personalizing another person's behavior toward them even when they should not.

The Self-Harm (BOR-S) subscale is not a measure of suicidal tendency, but is actually a measure of impulsivity. The construct of self-harm, as related to borderline personality disorder, includes recklessness of behavior and failure to consider consequences. Individuals with elevation on BOR-S ($T > 69$) act impulsively and do not think carefully through the potential consequences related to their behavior. Their impulsivity has caused problems for them socially, educationally, and occupationally. At significant elevations ($T > 85$), individuals' impulsivity is reckless and dangerous. They may be reckless in several domains, such as sex, spending money, or abusing substances, and they are at higher risk for directly self-harming behaviors, such as self-mutilation and suicidal behavior.

Frequent Code Types

The most frequent elevations that accompany elevation in BOR are on Anxiety-Related Disorders (see the "Anxiety-Related Disorders [ARD]" section above), Depression (see the "Depression [DEP]" section above), Schizophrenia (see the "Schizophrenia [SCZ]" section above), and Alcohol and Drug Problems (ALC and DRG, respectively). Borderline personality disorder has as one of its clinical features a pattern of reckless and impulsive behavior, and when BOR is elevated along with ALC or DRG, the use of alcohol or drugs is often part of this pattern of impulsive, often self-destructive behavior. Individuals with borderline personality disorder are emotionally labile and moody, and their interpersonal relationships are often characterized by volatility, conflict, and anger. The use of alcohol and drugs can easily exacerbate the already erratic shifts in mood and attitude experienced by the individuals elevated on BOR. It is often important to address the alcohol or drug use first in treatment, in order for these individuals to be able to benefit from the rest of treatment.

Treatment Implications

Elevations on BOR often coincide with individuals feeling they are in an acute crisis of some sort, and the motivation for treatment is often to help ameliorate that situation, despite the fact that clinicians may set goals to change the personal, often characterological symptoms that contribute more systematically to problems in the individuals' lives. As such, a major consideration for treatment is the potentially unaligned treatment goals between therapist and client. As elevations often mean an acute sensitivity, these individuals can feel easily judged and hurt, and attempts to acknowledge that they themselves play a significant (or sometimes any) role in their situational problems can lead to erratic anger or hurt feelings. Throughout the history of the mental health field, these symptoms have been associated with psychopathology that has intimidated clinicians. As more evidence emerges about how treatable borderline personality disorder is, clinicians are less scared or irritated by this clinical presentation. Still, the interpersonal aspects of building a therapeutic alliance with an individual elevated on BOR can be extremely challenging, and great care should be taken to align communication strategies with the preferred style and needs of the client.

Antisocial Features (ANT)

The ANT scale diverges from the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013)* more than the BOR scale does; while it includes behavioral elements of Antisocial Personality Disorder (which constitute all but one of the *DSM* criteria for the disorder) on the Antisocial Behaviors (ANT-A) subscale, the other elements included on the PAI related to antisocial features are actual personality traits, which the *DSM* does not capture well. Although the *DSM* includes psychopathy in the antisocial personality disorder diagnosis (i.e., individuals with psychopathy will be characterized as having antisocial personality disorder), many individuals who meet criteria for the disorder are not actually psychopathic. Thus, the ANT scale can characterize the behaviors (which can earn the diagnosis) as well as the personality traits generally expected in psychopathy. As with the BOR scale, examination of the subscales is critical in understanding the meaning of the ANT scale. Significant elevations on ANT ($T > 69$) represent individuals who are impulsive and potentially dangerous to themselves and especially to those around them. These individuals do not value the well-being of others and feel little if any remorse for their actions. Usually lacking in empathy, they have trouble valuing social norms and laws or caring about how breaking these norms and laws may affect others or society in general. These individuals seek novelty and stimulation, and they often take significant risks in order to find that stimulation. Marked elevation on ANT ($T > 82$) relates to a likelihood of meeting criteria for antisocial personality disorder, and when all three subscales are elevated, individuals display traits of psychopathy.

The Antisocial Behaviors (ANT-A) subscale focuses on the behaviors that constitute the bulk of the definition of antisocial personality disorder in the *DSM-5*, including history of conflict with authority figures, rule- and law-breaking behavior, and problems with following socially appropriate conventions. Individuals with elevations on ANT-A ($T > 69$) have had a history of rule-breaking behavior, difficulties with authority, and conventional behavior. It is important to note that this subscale asks both about current and past (during adolescence) patterns of behavior, as conduct problems typically start earlier than adulthood. At higher scores, the ANT-A reflects both a history of problematic behavior and current issues, including the possibility of physical aggression toward others and significant illegal behavior.

The Egocentricity (ANT-E) subscale focuses on an attitude of callousness and lack of regard for the feelings and well-being of others or society in general. The subscale includes a lack of empathy for others and a lack of significant, lasting remorse for negative behaviors toward them. Individuals with moderate elevations on ANT-E ($59 < T < 70$) are self-centered and calculated in their interpersonal relations, not allowing empathy or sympathy for others to impede them getting what they need, even if it is at the expense of others getting what they need. At significant elevations of ANT-E ($T > 69$), individuals have little regard for, responsibility for, or loyalty to those around them. They will easily exploit others for their own personal gain and not feel remorseful. Their primary goal is to get their own needs and wants met, regardless of what they need to do to achieve that. They are usually saliently callous and unfeeling toward others.

The Stimulus Seeking (ANT-S) subscale focuses on the strong urge for novelty and excitement and the willingness to take risks or act without considering the potential consequences in order to achieve it. It is important to note that this construct, in and of itself, may not be inherently negative, especially at moderate elevation. Many high-powered careers hinge on the ability to take major risks and gamble on their outcome. Individuals with moderate elevation on ANT-S ($59 < T < 70$), while potentially reckless, are not out of control in their seeking of novelty and excitement. They get bored easily, do not value routine or predictability, and would rather challenge convention. Significant elevations ($T > 69$) reveal potential danger to the individual him- or herself and to others. Their craving for novelty and excitement leads them to risky, dangerous behavior, and they may stir up trouble just for the thrill of it. These thrill-seekers look for a rush and are drawn to situations in which they can find one.

Frequent Code Types

The most common elevations that accompany ANT are related to Alcohol and Drug Problems (ALC and DRG). With both, the individual has had a history of problems related to acting out, most often involving substances, but also involving other problematic behavior. The use of substances is often instrumental in helping these individuals fulfill their need for excitement and novelty. Although their judgment is already questionable, alcohol and drugs help them make even more problematic, impulsive decisions in life. They are very likely to have interpersonal problems related to not taking others' feelings into account, and they are unlikely to give their interpersonal problems much thought. Often fun and exciting to be with, they seek thrills out wherever they go, and the use of alcohol and drugs help them do this.

Treatment Implications

Elevations on ANT often make the prospect of therapeutic treatment somewhat daunting. These individuals most often have very little motivation to change, especially with elevations on ANT-E, and the combination of impulsivity, poor frustration tolerance, and tendency to blame others makes forming trusting, close relationships quite rare for them. Their view of themselves is typically quite different from how others view them, and this, beyond the slow and steady building of a therapeutic alliance, can be a first point of intervention. Further, elevations on ANT can be addressed in treatment by helping the individuals understand the nature and consequences of these symptoms and traits, in order to exercise some behavioral control. Additional skills for self-monitoring and behavioral control may benefit these individuals as well.

Alcohol Problems (ALC)

The ALC scale uses both current and historical information to assess two different components of alcohol use-related problems. The first component is the clinical impact of using alcohol, such as social, educational, occupational, and legal consequences related to use. These consequences are important for identifying whether use is purely social, normative, and even perhaps adaptive or if it constitutes a diagnosable problem,

impairing functioning in some way. The second component assessed by the ALC scale is the presence of symptoms related to dependence, such as an inability to control use and withdrawal symptoms. This information is important in distinguishing between problem use and actual alcohol dependence, as these problems may warrant quite different types of treatment. It is important to note that the individual items that comprise the ALC scale ask directly about use of alcohol and consequences of that use. As such, it is easy for individuals to suppress this scale if they do not want to disclose their alcohol-related behaviors. Additionally, because much of the information gathered on ALC is historical in nature, elevations on this scale (as well as profiles without elevations but with elevation in ALC Est, as discussed earlier in the “Positive Impression [PIM, DEF, CDF, ALC Est, DRG Est]” section) should trigger further assessment of alcohol-related problems. Significant elevations on ALC ($T > 69$) represent individuals who would now or in the past at least qualify as struggling with alcohol abuse, with their use affecting their functioning in some way. They likely have (or have had) problems related to interpersonal functioning, job performance, or even health concerns all related to their alcohol use. At marked levels of elevation of ALC ($T > 83$), there is an increased likelihood that the individual is currently experiencing alcohol dependence, including an inability to stop drinking, needing more alcohol to reach the same levels of intoxication, and suffering withdrawal symptoms when denied alcohol. Alcohol use is negatively affecting the individual’s life, likely across multiple domains, and he or she feels guilty and ashamed of the alcohol use and its consequences.

Frequent Code Types

Alcohol problems co-occur with many different clusters of symptoms, including Somatic Complaints (see the “Somatic Complaints [SOM]” section above), Anxiety (see the “Anxiety [ANX]” section above), Depression (see the “Depression [DEP]” section above), Schizophrenia (see the “Schizophrenia [SCZ]” section above), Borderline Features (see the “Borderline Features [BOR]” section above), and Antisocial Features (see the “Antisocial Features [ANT]” section above). Additionally, ALC is very frequently seen elevated concurrently with Drug Problems (DRG). Individuals with elevations on both ALC and DRG have a history of using alcohol and other drugs, and this use has caused significant impairment in their lives related to their interpersonal or work functioning. These codes can also be elevated together in combination with the other code types that frequently accompany either one separately.

Treatment Implications

Going beyond just a measure of past and current alcohol use and abuse (and potential dependency), elevations on ALC reflect at least some acknowledgement that their alcohol use has caused problems in their lives. As such, these elevations may serve as a good prognostic indicator, even though alcohol problems themselves can be difficult to treat. Once it has been ascertained whether the alcohol use and its associated problems are entirely in the past or continue currently, targeted treatment can be employed to address the alcohol use. It is often wise to address alcohol use as early in treatment as possible, even before addressing some other problems, as the use and abuse of alcohol can mask, exacerbate, or even disguise other psychological problems.

Drug Problems (DRG)

Like the ALC scale, the DRG scale uses both current and historical information to assess two different components of problems related to drug use. The first component is the clinical impact of using drugs, such as social, educational, occupational, and legal consequences related to use. The second component is the presence of symptoms related to dependence, such as an inability to control use and withdrawal symptoms. This information is important in distinguishing between problem use and actual drug dependence, as these aspects of using drugs may warrant quite different types of treatment. It is again important to note that the individual items that comprise the DRG scale ask directly about use of drugs and consequences of that use, again making denial of problems quite easy for respondents. Additionally, like ALC, because much of the information gathered on DRG is historical in nature, elevations on this scale (as well as profiles without elevations but with elevation in DRG Est, as discussed earlier in the “Positive Impression [PIM, DEF, CDF, ALC Est, DRG Est]” section) should trigger further assessment of drug-related problems. Significant elevations on DRG ($T > 69$) represent individuals who would now or in the past at least qualify as struggling with drug abuse, with their use affecting their functioning in some way. They likely have problems related to interpersonal functioning, job performance, or even health concerns, all related to their drug use. At marked levels of elevation of DRG ($T > 79$), there is an increased likelihood that the individual is currently experiencing drug dependence, including an inability to stop using and suffering withdrawal symptoms when not actively using. Drug use is negatively affecting the individual’s life, likely across multiple domains, and he or she feels guilty and ashamed of the drug use and its consequences.

Frequent Code Types

As discussed earlier, DRG elevations are often found with multiple other clinical scales, including Somatic Complaints (see the “Somatic Complaints [SOM]” section above), Anxiety (see the “Anxiety [ANX]” section above), Depression (see the “Depression [DEP]” section above), Schizophrenia (see the “Schizophrenia [SCZ]” section above), Borderline Features (see the “Borderline Features [BOR]” section above), Antisocial Features (see the “Antisocial Features [ANT]” section above), and Alcohol Problems (see the “Alcohol Problems [ALC]” section above).

Treatment Implications

Similar to ALC, elevations on DRG can signify some insight into the fact that use of drugs has caused significant problems in an individual’s life, and the use of drugs, if determined to be current, should be addressed early in treatment. Elevations in DRG may signify a readiness for change, whereas low scores on DRG accompanied by high DRG Est scores may reflect significant denial or an unwillingness to change. When elevated, and when further evaluation clarifies the exact nature and severity of the problem, targeted drug treatment may be warranted, possibly including detoxification and intensive ongoing rehabilitation. When it is determined that elevations on DRG relate entirely to the past, treatment can acknowledge both the fact that drugs used to play

a significant role in the individual coping with and interacting with the world and the strength and accomplishment that the individual changed that pattern.

TREATMENT SCALES

Aggression (AGG)

The AGG scale provides a global evaluation of aggressive attitudes and behaviors, which relate to many different forms of psychopathology. The scale is considered a treatment scale because of the potential centrality of this social attitude and behavior on the therapeutic relationship. The subscales are extremely important in identifying how the individual is likely to manifest his or her aggression, whether in verbal or physical ways or in passive-aggressive ways. Moderate elevation on AGG ($59 < T < 71$) represents individuals who are impatient, irritable, and easy to anger. These individuals may be seen as “snippy” toward others, though they are unlikely to lash out significantly. Significant elevations on AGG ($T > 70$) represent individuals who are both generally angry and hostile and fully prepared to exhibit their anger to others. These individuals maintain a consistently hostile attitude and will lash out at others when they feel angry. Marked elevation on AGG ($T > 82$) relates to true anger management problems and the potential for dangerous aggression toward others. These individuals tend to be explosive with their anger, and their interpersonal relationships and job performance can suffer because of it.

The Aggressive Attitude (AGG-A) subscale focuses not on aggressive behavior, but rather on the traits and beliefs that underlie anger and aggression, such as the inability to control one's temper and the belief in the utility of aggressive behavior. Individuals with moderate elevations on AGG-A ($59 < T < 71$) are irritable and somewhat hot-tempered. They are easily provoked to anger by others, and they likely have angry outbursts in situations that would not warrant anger from others. Significant elevations on AGG-A ($T > 70$) suggest an individual with significant anger issues, losing his or her temper easily even when minimally provoked. When AGG-A is elevated, it is important to note which, if either, of the other AGG subscales is elevated, to determine how the individual is likely to express his or her anger. If neither of the behavioral subscales of AGG is elevated, it is likely that the individual is actively suppressing anger or may exhibit it in a more passive-aggressive manner.

The Verbal Aggression (AGG-V) subscale focuses on how likely an individual is to express his or her anger by verbal means, ranging from mild sarcasm to overt, angry yelling and abusive language. Individuals with moderate elevations on AGG-V ($59 < T < 71$) are unafraid to be assertive, though at times they may cross a boundary into aggressive or pushy language. They have no problem with verbal confrontation. At significant elevations of AGG-V ($T > 70$), individuals are verbally hostile and abusive toward others, channeling their anger into tirades and cutting, bitter language. Going far beyond assertive, their use of verbal aggression violates others and does not respect their boundaries, situations, or feelings.

The Physical Aggression (AGG-P) subscale assesses both current attitude toward and historical use of physical aggression toward others. Similar to some other scales,

historical information can be important, as problems with containing anger and channeling it physically rarely surface for the first time in adulthood. Individuals with moderate elevation on AGG-P ($59 < T < 71$) have had some history of physical aggression, including breaking objects or even physical violence toward others, but they manage to maintain control over their anger and aggression most of the time. Significant elevations ($T > 70$) are related to a greater frequency of losing one's temper and consequently destroying property or threatening violence toward others. These physical outbursts are often sudden and may seem unprovoked to an outside observer, and these individuals are dangerous to others and intimidate them easily.

Treatment Implications

The implications for treatment related to AGG focus squarely on the therapeutic relationship. Both low and high scores on AGG have implications for the way the client is likely to relate to the therapist and to treatment in general, and the AGG-A subscale has particular implications for treatment. High scores on AGG relate to clients who are going to be resistant, combative, and potentially explosive toward the therapist. For these individuals, clinicians should be careful about the therapeutic alliance, while at the same time addressing anger management problems outright, including the possible use of relaxation techniques and training for problem solving and other social skills. When AGG is extremely low, clients tend to be passive and submissive, and it is highly possible that they will be acquiescent with therapy, even when and if they do not actually agree with the therapist. Assertiveness training may be important for these individuals.

When AGG-A is particularly elevated, especially in the absence of elevation of AGG-V and AGG-P, the clinician should be especially aware of underlying anger that is either suppressed and unacknowledged or may come out in passive-aggressive ways. These individuals may participate actively and agree with the clinician in sessions, but then they may undermine these efforts outside of sessions. Helping these individuals find healthy and appropriate ways to express their anger, disagreement, or judgments within the therapeutic relationship may be a specific goal of treatment.

Suicidal Ideation (SUI)

The SUI scale evaluates a spectrum of warning signs related to suicidality, from feelings of hopelessness about life in general, to vague thoughts related to death and dying, all the way to specific plan and intent to commit suicide. Similar to the ALC and DRG scales, SUI asks specifically and directly about suicidal thoughts and plans, and thus it is easy for respondents to suppress this scale simply by denying the presence of these thoughts. Other indicators on the PAI, such as the DEP, the Stress (STR) scale, and the Nonsupport (NON) scale, may indicate that the clinician should evaluate further into suicidal potential, even when SUI is low. Moderate elevations on SUI ($59 < T < 70$) tend to reflect pessimism and fleeting, passive thoughts about suicide, which is common in clinical settings. Individuals who are depressed, for example, even if they are not actively suicidal, often entertain thoughts about what would happen if they were dead. As this is a high-stakes area of assessment, even these moderate elevations should be

highlighted and followed up on. Significant elevations on SUI ($T > 69$) signify active suicidal ideation, which must be followed up on. Although elevations may constitute a cry for help and may not reflect true intention to kill oneself, the clinician should always err on the side of addressing potential suicidality when it emerges. Individuals with imminent, clear, and specific plans to harm themselves typically score markedly high on SUI ($T > 83$), and those who score extremely high ($T > 100$) are preoccupied with suicidal ideation and have likely taken some steps toward an attempt.

Treatment Implications

Rather than discuss the treatment strategies for suicidality, which are quite complex and detailed in the literature, it is most important to note that elevations on SUI should trigger two processes. First, other scale elevations should be considered for their potential relationship with suicidality. For example, DEP-C, BOR-S, and DRG provide information on hopelessness, impulsivity, and substance-related behavioral disinhibition, respectively. Elevations in all of these, along with elevation on SUI, suggest significant potential for suicidality. The second process that should be triggered with elevations on SUI is a more comprehensive, targeted suicidal assessment, which can ascertain details that are not captured on the PAI, such as overtly stated intent and specific plans for committing suicide. The SUI and the PAI in general are not meant to provide a thorough assessment of suicidality, and they should be used only as a red flag to trigger further evaluation.

Stress (STR)

The STR scale evaluates the current level of daily life stressors being experienced by the respondent, including family, work, and financial stressors and other major life changes that, regardless of their actual nature, place stress on the individual. The scale does not provide specific details about the nature or content of the current stress, but rather focuses on the quantity and impact of stresses on the individual. As such, elevations on STR should trigger further evaluation by the clinician in order to determine the exact nature of the stressors in the individual's current life. While some elevation represents some stress, significant elevation on STR ($T > 69$) signifies that the current level of stress in the individual's life is significantly impacting him or her in a negative way, likely too much for the individual to handle on his or her own. Marked elevation on STR ($T > 84$) relates to individuals who feel constantly surrounded by crises, unable to manage and control what is happening in their life. These individuals are especially prone to psychopathology (of many different types) and may suffer from "breakdowns" related to their heightened stress levels.

Treatment Implications

Typically, the higher the score on STR, the higher the motivation level for individuals to get help, including treatment. As STR is a measure of specific stressors currently pressuring the individual, it is important for treatment to include an exploration of what those stressors are. Very often those elevated on STR will respond well to problem-solving strategies and support in ameliorating their current situations,

including learning skills to better cope with pressure and stress and identifying behaviors that are contributing to the stressful situations in their lives. It is important to note that individuals with elevations on STR, regardless of what other scales or subscales are elevated, may have a difficult time focusing on larger or deeper issues in treatment while they still perceive that there are major stressors in their lives. That is, treatment may benefit from focusing on the stressful situations first, then moving on to deeper or more global work.

Nonsupport (NON)

The NON scale evaluates the respondent's perception of how well supported he or she is by others in his or her life. The level of support relates both to the quantity of support (how available support is to the individual) and its quality (including the nature of interactions with family and friends). Low scores on this scale relate to high perceived support, while high scores relate to a perception that the individual has too little social support in his or her environment. While some elevation represents the likelihood of few close supports or dissatisfaction with the support network, significant elevation on NON ($T > 69$) signifies that the individual currently feels that there is very limited support offered by friends or family. Marked elevation on NON ($T > 87$) signifies that individuals feel that have little or no social support during significant life events. They are not close with their families and have few friends they feel they can rely on. They tend to be critical of others, who they feel are unsupportive and cold, but they are also critical of themselves, having few solid relationships.

Treatment Implications

When NON is elevated, individuals are likely actively seeking support, and treatment can and should fill this need. It is important for these individuals to feel that the therapeutic context is supportive and caring, as long as the elevated NON is not related to elevations in PAR or SCZ, which may reflect that the individuals have little support but also want little support. There are cases in which elevation on NON reflects specific turmoil in primary relationships, such as family or partner discord. In these cases, couple or family therapy may be considered as a primary form of treatment, in order to address problematic interpersonal patterns in these primary attachment relationships.

Treatment Rejection (RXR)

The RXR scale was developed to evaluate the degree to which the respondent is willing to admit that he or she needs to change or grow in some way, willing to accept responsibility for problems in his or her life, and willing to engage in a treatment process to better him- or herself. The T scores for this scale are calculated in comparison to a general population sample, in which many subjects would not have had overt, bothersome mental health issues. As such, for an individual who is presenting for an evaluation with some problems, even average scores may be somewhat problematic; thus, the threshold for interpretation of this scale is slightly lower than the other scales. Low scores on RXR ($T < 41$) suggest an individual who understands that he or she has problems,

plays a role in those problems, and is willing to ask for and receive help for those problems. Extremely low scores ($T < 20$) generally reflect individuals who are significantly and acutely overwhelmed and need assistance immediately. An individual who scores in this range would be expected to score in a significant or even markedly high range on one or more of the clinical scales. Moderate elevation on RXR ($49 < T < 60$) reflects general satisfaction with who the individual is and what is going on in his or her life at the moment, with little felt need for change. Significant elevation ($T > 59$) represents individuals who do not easily admit to problems or faults and will be resistant to change or even engagement in a therapeutic process. These individuals often feel that problems in their lives have little to do with their own behaviors, that they stem primarily from other people. As such, they tend to be quite resistant to changing the way they do things, including the way they think or behave.

Treatment Implications

True to its title, individuals who have elevated RXR scales are difficult to engage in a therapeutic situation. As they do not necessarily feel the need to change or grow in any particular way, even convincing them that treatment is a good use of time or money may be a challenge. Strategies that search for any ambivalence may be useful for these individuals, as the clinician can capitalize on any thoughts or feelings related to the individual even considering the need to change. Additional strategies may include using a nondirective, collaborative approach and/or paradoxical interventions. It would also be useful to tailor intervention according to stage of change and use motivational interviewing techniques.

INTERPERSONAL SCALES

Dominance (DOM)

The DOM scale on the PAI reflects the degree to which an individual relates to others in a balanced way, respecting his or her own thoughts, feelings, and opinions as well as those of others, versus relating to others in an imbalanced way, either dominating and controlling interactions or submitting to others passively. The scale reflects a bipolar construct, such that extreme scores in either direction are potentially problematic. Individuals who relate to others in flexible ways, generally both autonomous and respectful, score in the average range ($44 < T < 60$). Moderately low scores on DOM ($34 < T < 45$) reflect individuals who are generally passive and reticent to assert themselves, uncomfortable with being the center of attention or needing to make decisions for groups of people. At extremely low scores on DOM ($T < 35$), individuals lack self-confidence in social situations and interpersonal relationships, as well as the ability to get their needs met adequately in interactions with others. They are especially vulnerable to being exploited or taken advantage of by others. Moderately elevated scores on DOM ($59 < T < 70$), conversely, relate to socially confident, assertive individuals who are unafraid to make their needs and wants known and even control situations and others to get them met. These individuals are not necessarily overbearing or pushy, but they tend to be natural leaders and to prefer situations in which they can act as such.

Significant elevations on DOM ($T > 69$) reflect a tendency to be overbearing, dominating, and intolerant of disagreement from others. These individuals are driven to always appear competent and even expert, and they strongly value being treated with admiration.

Treatment Implications

Elevations on DOM can signify a client who may potentially engage in power struggles with the therapist. That is, these individuals prefer to be in the more prominent, commanding role in relationships, and they may find it difficult to cede power and influence to the therapist. However they choose to address the dynamic, clinicians should be careful not to take on a submissive or passive role in treatment, simply because this is what the client prefers. Clinicians will need to balance the therapeutic alliance (which will pull for submission) with treatment interventions (which will require the clinician to take on a more active role). Often a collaborative, nondirective approach may be effective. Similarly, when an individual enters treatment with a very low DOM score, his or her passivity and submissiveness will likely pull for the clinician to be more assertive and domineering. Again, the clinician must strike a balance between building and maintaining a positive relationship (which will pull for directiveness on the part of the therapist) and understanding that the client may need to learn to be more assertive and self-promoting.

Warmth (WRM)

The WRM scale reflects the level of empathy, sociability, and affection that an individual believes he or she has toward others. Like the DOM scale, the WRM scale reflects a bipolar construct, with both low and high scores being interpretable. Individuals who score in the average range ($44 < T < 60$) are flexible in the way they approach interpersonal interaction, able to tolerate closeness and affection within significant relationships, but also able to maintain distance when necessary. Moderately low scores on WRM ($34 < T < 45$) reflect individuals who are generally somewhat distant in relationships, not valuing deep, close, emotional bonds with others. They are practical and independent, rarely swayed by others, and they may be seen as somewhat cold in social interactions. At extremely low scores on WRM ($T < 35$), individuals are uninterested in forming deep connections with others, often unable to show emotion within interpersonal interactions. They are cold and impassive, not showing affection or empathy easily with others. Moderately elevated scores on WRM ($59 < T < 70$), conversely, relate to individuals who are warm, empathic, and value deep, significant relationships. These individuals place value on getting along with each other and are generally friendly and sympathetic. They may be uneasy with confrontation, but they can assert themselves when needed, though usually in a very respectful way. Significant elevations on WRM ($T > 69$) reflect a tendency to be overly concerned with being liked and accepted by others, often at the expense of getting one's own needs met. Individuals in this range can be too caring and too giving, and they will avoid conflict with others to an extreme degree. These individuals are particularly vulnerable to being exploited for their friendliness and sympathy.

Treatment Implications

The major treatment implication related to WRM relates to how clinicians are likely to react to and feel about the client. Clients high in WRM will elicit positive, warm, caring reactions from their therapists overall, while clients low in WRM will likely receive less positive evaluation and emotional connection. Clinicians can use the WRM scale score to anticipate these interpersonal dynamics, as well as to think about how they may play out in the world outside of therapy. Often it is in the best interest of treatment for clinicians to understand these interpersonal dynamics and work hard not to be too heavily influenced by them. Obviously, negative reactions to the interpersonal dynamics with individuals low on WRM can be counterproductive to treatment. However, even the positive reactions to those high on WRM can create blind spots for therapists, such that they do not notice negative aspects of the client. This halo effect can be just as detrimental to treatment as negative feelings and impressions of the client on the part of the clinician.

ADDITIONAL CLUSTERS

Self-Concept

Several variables can be used in conjunction with one another to help practitioners understand how individuals think and feel about themselves. The three primary variables used to evaluate self-concept are Grandiosity (MAN-G), Cognitive Depression (DEP-C), and Identity Problems (BOR-I). When looked at together, these three variables show something about how individuals think and feel about themselves and how stable this perception is. MAN-G reflects the affective component of self-concept, self-esteem. At markedly high levels ($T > 69$), MAN-G represents efforts to respond to actual low self-esteem; that is, when MAN-G is extremely high, although the individual is reporting high self-esteem, it is grandiose to a level that reflects defensive functioning, in reaction to feeling that the individual is not good enough. Although the defensive style and reaction is different, low scores on MAN-G ($T < 41$) reflect a similar underlying fragility of self-esteem. These individuals do not like themselves and would prefer to be different from who they are. At average levels ($40 < T < 70$), self-esteem is generally intact, with higher scores reflecting healthier and stronger positive feelings about the self.

The thought component of self-concept, self-efficacy, is reflected most directly in the DEP-C subscale. High scores on DEP-C ($T > 74$) reflect beliefs that the individual is ineffective and unable to control his or her environment, change his or her situation, or meet his or her needs. Depressive thoughts center on helplessness and worthlessness, beliefs that even if the individuals have specific goals in life, they will not be able to achieve them. Average scores on DEP-C ($50 < T < 75$) reflect an occasional tendency to be self-critical. In general, thoughts about the self may fluctuate, from generally positive and confident during periods of ease to questioning competence during times of stress and uncertainty. Low scores on DEP-C ($T < 51$) reflect a generally resilient and self-assured attitude about the self, with the person confident that he or she can effect change in his or her life as needed. When combined with low or high MAN-G, high DEP-C relates to a generally consistent and negative self-concept.

MAN-G and DEP-C evaluate the content of self-concept; BOR-I relates to the consistency and stability of those thoughts and feelings about the self. High scores on BOR-I ($T > 74$) relate to self-concepts that are poorly defined and understood by the individual him- or herself. While the MAN-G and DEP-C may indicate how the individual is currently feeling and thinking about him- or herself, with high BOR-I, this self-evaluation is likely to fluctuate and change significantly in different circumstances and contexts. External events will strongly influence how the individual thinks and feels about him- or herself. Average scores on BOR-I ($50 < T < 75$) reflects some consistency in the way the individual evaluates him- or herself, but with some reactivity to external events and circumstances. Individuals in this range may have fluctuating feelings and thoughts about themselves depending on the relative failures and successes in their lives. Low scores on BOR-I ($T < 51$) reflect consistency and stability in the way individuals view themselves. Not inherently good or bad, scores in this range can be paired with relatively positive or negative MAN-G and DEP-C scores, such that the negative or positive thoughts and feelings are generally consistent.

When evaluating the three subscales together, the clinician can begin to understand the general self-concept of the individual, how consistent it is from one situation to the next, and how generally positive or negative it is. Intervention decisions can be made based on the different configurations of the three scores. For example, an individual with a low MAN-G, high DEP-C, and low BOR-I would be someone who consistently views him- or herself in a negative light. He or she has a clear view on him- or herself, and that view comes with low self-esteem and low self-efficacy. Intervention may focus on the validity of these self-evaluations, potentially challenging the overly negative view of the self. If the same person had emerged with a low MAN-G, high DEP-C, and *high* BOR-I, while he or she is currently feeling negatively about him- or herself and questioning his or her abilities, focus of intervention may be better placed on the fact that his or her sense of who he or she is is poorly established and defined. Efforts to intervene on the negative thoughts or feelings about the self may appear to be successful, but improvements may represent the general lability of self-concept, rather than a more pervasive improvement. Intervention may better be targeted on helping the individual create a clear and stable (hopefully positive) sense of who he or she is.

Interpersonal Style

The two interpersonal scales, DOM and WRM, can be evaluated together to represent the general interpersonal style of the individual being evaluated. Although interpersonal style is often more nuanced and complicated than this, in general, four configurations of DOM and WRM together represent four different types of interpersonal styles. Those high on both DOM and WRM ($T > 55$) are typically extraverted and socially comfortable and effective. These individuals generally have many friends and parent effectively, taking control of situations when needed, but doing so in a truly caring, compassionate way. When these individuals have extreme elevations on DOM and WRM ($T > 65$), they likely have a very strong need to be noticed and liked. These individuals may exhibit dramatic and exaggerated behaviors, all toward the end of meeting their own needs to be appreciated.

When DOM is high ($T > 55$) and WRM is low ($T < 45$), individuals likely view interpersonal interactions and relationships as means to ends. They have their own agendas and needs to be met, and they are unsentimental about doing what needs to be done in order to meet them. When these scales are exaggerated, such that DOM is very high ($T > 65$) and WRM is very low ($T < 35$), individuals are generally exploitative and manipulative. They are often egocentric, and they see no value in having relationships for the sake of having relationships; relationships exist to meet their own needs. These individuals are controlling and unsympathetic toward others.

When both DOM and WRM are low ($T < 45$), individuals are introverted, often isolated, and distanced in their relationships with others. These individuals often have very little need for and see very little value in close relationships with others, and they will rarely go out of their way to appear friendly or put on what they feel are unnecessary social graces. At exaggerated levels, with very low DOM and WRM ($T < 35$), individuals are markedly uncomfortable in social situations and make efforts to avoid them, being passive or submissive when forced to interact with others in order not to engage socially more than they absolutely have to. They have some paranoia or resentment, and scales tapping these constructs should be examined with this pattern of interpersonal scales.

When DOM is low ($T < 45$) and WRM is high ($T > 55$), individuals tend to be cooperative and friendly with others, preferring to get along and maintain relationships rather than be a leader or question authority. They are more comfortable in a followership role, and they tend to be extremely supportive and genuine in their interactions. With exaggerations on these scales, such that DOM is very low ($T < 35$) and WRM is very high ($T > 65$), an individual likely fears rejection and abandonment, preferring to cede any personal authority or preference to others and often not getting his or her needs met. These individuals are easily taken advantage of, as their dependency needs overshadow any ability to be assertive. They may feel ineffective and can easily become overwhelmed even by minimally stressful situations.

It is important to note that these four configurations become more nuanced and complex when average scores on each the DOM and WRM scale are considered. For example, an individual who scores within the average range on both DOM and WRM ($44 < T < 56$) tends to be more balanced in his or her relationships and interactions with others, autonomous and adaptable to the needs of each situation. The clinician should consider not only the extreme versions of these configurations, but also the implications of other potential patterns.

Environmental Perception

Similar to the interpersonal style cluster, two scales, NON and STR, can be evaluated together to represent how the individual being evaluated perceives his or her current environment. Again, although it is often more nuanced and complicated, in general, four configurations of NON and STR together represent four different general views of current context. Those average or low on both NON and STR ($T < 60$) feel that they have stable, reliable support in their life and do not feel overwhelmed by stressors or unexpected pressure. These individuals typically have good relationships, people on whom they can rely, and do not feel they have overwhelming demands being placed on

them very frequently in their lives. This pattern is associated with positive prognosis in treatment, as these individuals tend to have enough social support to help them handle their problems. Extremely low scores on both NON and STR ($T < 44$) are an excellent sign of potential for personal improvement, though any problems that do exist are likely to cause guilt, as these individuals have no environmental stressors or lack of support to blame their issues on.

When NON is high ($T > 59$) and STR is average or low ($T < 60$), individuals feel that other people offer little comfort, support, or useful help in times of need. However, these individuals do not currently feel that they have a real need for support, as they do not perceive many major stressors in their life. Although they may not be close with friends and family, and sometimes *because* they are not close with friends and family, their lives are stable, predictable, and without major unexpected pressures. When NON is very high ($T > 84$), these individuals not only do not feel others are particularly helpful, but they view others as abandoning, rejecting, and cold. They have no faith that others can be helpful or genuinely caring, and they often hold disdain for others for this lack of sympathy and support. It is very common for this pattern to be accompanied by elevations in BOR-N and PAR.

When NON is average or low ($T < 60$) and STR is high ($T > 59$), individuals are experiencing significant situational stress but feel that they have a good deal of support from others around them to help them cope effectively with it. Elevations on clinical scales, especially DEP and ANX, are likely related to this situational pressure, and it is very likely that individuals with this pattern will improve (and benefit from treatment), given their good social support. When STR is very high ($T > 84$), individuals are currently so overwhelmed by whatever is going on in their environment that they are experiencing significant impairment. Although they feel they have support, it is highly possible that they are not proactively reaching out for help, and they should be encouraged to do so.

When both NON and STR are high ($T > 59$), individuals are experiencing notable stress, often across multiple areas of life, and likely especially within the realm of interpersonal relationships. They are reporting that others are not supportive, and they likely have limited close relationships, with the relationships they do have likely characterized by conflict. While it is likely that not all of the individual's problems are related to interpersonal relationships, relationships are problematic and should be targeted in treatment as at least one goal. At exaggerated levels, with very high NON and STR ($T > 84$), individuals are reporting severe instability in their life situation and a complete lack of social resources to which to turn for help. Relationships are a key stressor in these individuals' lives, and elevations on BOR-N and PAR may be present as well. Resentment and bitterness may be present, and it is highly possible that the chaos in the individuals' lives is not simply situational, but the individuals likely play a significant role in maintaining more chronic turmoil, with personality problems potentially present. Accordingly, long-term treatment is warranted.

Again, it is important to note that these four configurations become more nuanced and complex when a more continuous-scale interpretation of scores on each the NON and STR scale is considered. For example, an individual who scores within the moderately elevated range on STR ($59 < T < 70$) but the low range on NON ($T < 44$) is experiencing some stressors, perhaps even a normal amount of stressors if there are

elevations on any of the clinical scales, but feels he or she has an abundance of support to help him or her cope with the stressors. The clinician should consider not only the extreme versions of these configurations, but also the implications of other potential patterns.

Potential for Dangerousness

Although the PAI has scales directly related to the overt reporting of potential dangerousness, most notably SUI for self-harm and AGG for the potential to harm others, the measure also includes two configural indices to evaluate the potential for these dangerous behaviors that do not require the direct reporting of these tendencies on the part of the individual being evaluated. Morey (2007) developed the Suicide Potential Index (SPI) and the Violence Potential Index (VPI) as measures of personality and clinical traits associated with suicidality and violence, respectively, in order to try to distinguish, in a less obvious way, those individuals with dangerousness potential from those without. Both indices include 20 risk-related features, set up as variables with decisional cutoffs. The clinician compares the findings on each scale and subscale listed with the cutoff scores listed in each index's table to determine whether or not a criterion is met; then the clinician makes decisions for dangerousness potential based on how many feature criteria are met.

The SPI evaluates variables within the PAI that are risks associated with an elevated potential for suicidality, above and beyond the overt reporting of suicidality measured by the SUI. Included in these variables are information related to heightened negative emotion, low positive emotion, volatility and moodiness, and problems with impulse control (see Table 8.2). For example, some of the variables that measure high negative affect include cognitive anxiety (ANX-C, which includes worry and rumination), cognitive depression (DEP-C, which includes hopelessness), and affective depression (DEP-A, which includes sadness and distress). A raw score of 13 (when an individual meets 13 out of 20 of the criteria) is equal to a *T* score of 81, which is not only dramatically above the mean for the general community sample but is in fact 1 standard deviation above the mean for the PAI clinical sample. Scores in this range reflect a heightened possibility that an individual is in a current mental space that renders him or her to possibly be considering suicide. This level of distress is significant, and further evaluation and intervention should be undertaken to understand the seriousness of the individual's current circumstances.

The VPI evaluates variables within the PAI that are risks associated with an elevated potential for violence against others, above and beyond the overt self-report of aggressive tendencies that are measured by AGG (see Table 8.2). Included in these variables are measures of antisocial behavior in general (ANT-A), mood lability (BOR-A), agitation (MAN-A), and paranoia (PAR-H, which includes suspiciousness, and PAR-P, which includes a sense of being persecuted by others), among other traits. Additional factors relate to disinhibition of behavior, both by substance use (ALC and DRG) and more characterological (BOR-S, which measures impulsivity). The combination of negative, angry, resentful, or unpredictable internal processes and disinhibition of behavior are major risk factors for potential for dangerousness toward others. A raw score of 9 (when an individual meets 9 out of 20 of the criteria) is equal to a *T* score of

Table 8.2 Validity, Clinical, Treatment, and Interpersonal Personality Assessment Inventory Scales

Suicide Potential Index (SPI)		Violence Potential Index (VPI)	
Scale/Subscale	<i>T</i> score cutoff for inclusion	Scale/Subscale	<i>T</i> score cutoff for inclusion
ANX-C	>60	AGG-P minus AGG-V	≥15
DEP-A	>65	AGG minus SUI	≥10
DEP-P	>60	DOM minus WRM	≥10
SCZ-T	>60	ARD-T minus ARD-P	≥15
ARD-O	>55	ANT-A	>70
ALC	>60	ANT-E	>60
ANX-P	>60	ANT-S	>60
MAN-A	>55	BOR-A	>70
NON	>60	BOR-N	>70
DRG	>60	BOR-S	>70
BOR-N	>65	MAN-A	>60
STR	>65	MAN-G	>60
BOR-S	>60	NIM	>70
AGG-P minus AGG-V	>10	PAR-H	>70
DEP-C	>65	PAR-P	>70
PAR-H	>60	SCZ-P	>70
WRM	<45	SCZ-S	>70
MAN-G	<45	ALC	>70
BOR-A	>65	DRG	>70
SOM-H	>55	NON	>70

Source: Adapted and reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, FL 33549 from the Personality Assessment Inventory by Leslie C. Morey, PhD and PAR Staff, Copyright 1991, 1992, 2005, 2007. Further reproduction is prohibited without permission of PAR.

84 when compared to the general population community sample, and this raw score is 1 standard deviation above the mean of the clinical sample. A raw score of 17 is equivalent to a *T* score of 121 and is 2 standard deviations above the mean of the clinical sample. These elevations represent risk for violent behavior and should be evaluated in conjunction with the AGG scale and subscales, as well as with a detailed history of aggressive or violent behavior.

CRITICAL ITEMS

Twenty-seven items are considered critical items on the PAI, such that endorsement of them signals the clinician that there is a need for immediate follow-up. These items

were endorsed relatively infrequently in the normative samples, and they are clinically considered potentially high risk. They may signify a current crisis situation or the potential for imminent negative outcomes. Some examples of these items include endorsing having made plans to kill oneself (potential for self-harm), endorsing that others are afraid of one's temper (potential for harming others), and being troubled by and reliving bad things that have happened (traumatic stressors), among others. Clinicians should examine these individual items carefully and determine if follow-up is warranted.

RECOMMENDED READING

- Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (2003). *Essentials of PAI assessment*. Hoboken, NJ: Wiley.
- Morey, L. C. (2007). *Personality Assessment Inventory professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (1998). Teaching and learning the Personality Assessment Inventory (PAI). In L. Handler & M. Hilsenroth (Eds.), *Teaching and learning psychological assessment* (pp. 191–214). Mahwah, NJ: Erlbaum.

MILLON CLINICAL MULTIAXIAL INVENTORY

The Millon Clinical Multiaxial Inventory (MCMI) is a standardized, self-report questionnaire that assesses a wide range of information related to a client's personality, emotional adjustment, and attitude toward taking tests. It has been designed for adults (18 years and older) who have a minimum of an eighth-grade reading level. The MCMI is one of the few self-report tests that focus on both personality disorders and the symptoms frequently associated with these disorders. Originally developed in 1977 (Millon, 1977), it has since been through three revisions (MCMI-II; Millon, 1987; MCMI-III; Millon, 1994, 1997; MCMI-IV; Millon, Grossman, & Millon, 2015). Since its original publication, it has stimulated more than 1,000 published papers on or about it and has become one of the more frequently used tests in clinical practice (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Camara, Nathan, & Puente, 2000; C. Piotrowski & Zalewski, 1993). Indeed, it is one of the few tests that has risen through the ranks of test usage over the past 40 years. Among objective personality tests for clinical trainees to be familiar with, the MCMI was ranked by directors of clinical training programs second only to the MMPI/MMPI-2 in importance (C. Piotrowski & Zalewski, 1993), though this information is quite dated at this point. Its popularity is further supported by its use in several different countries and its translation into a number of different languages.

The current version, the MCMI-IV, is composed of 195 items that are scored to produce 30 scales divided into the following categories: Validity and Modifying Indices, Clinical Personality Patterns, Severe Personality Pathology, Clinical Syndromes, and Severe Syndromes (see Table 9.1). The scales, along with the items that comprise the scales, are closely aligned to both Millon's theory of personality and the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* and *International Classification of Disorders (ICD)* (Millon, 2011). For example, an item endorsing a person's belief in his or her own superiority would be part of the Narcissistic scale, because the content clearly relates to components of Millon's and the *DSM*'s conceptualization of the narcissistic personality. Many of the scales have both theoretical and item overlap—an important fact to keep in mind when conceptualizing the client and interpreting the scales. Thus, an elevation on both the Antisocial and Sadistic scales would reflect a person who has sadistic features along with legal difficulties and impulsiveness, and who is interpersonally exploitive. Similarly, a person scoring high on the Antisocial scale might have a corresponding elevation on the Alcohol Dependence scale. The corresponding elevations on conceptually related scales allow for a more complete understanding of the client.

In some ways, the MCMI is an alternative or even a competitor to the MMPI. Both instruments cover a wide range of adult pathology that assesses both long-standing

Table 9.1 MCMI-IV Scale Categories, Abbreviations, Number of Items, and Reliabilities

Scale category/name	Scale abbreviation	Scale number	No. of items	Alpha
Validity/Modifying Indices				
Invalidity		V	3	
Inconsistency		W	50	
Disclosure		X	121	
Desirability		Y	24	
Debasement		Z	30	
Clinical Personality Patterns				
Schizoid	AASchd	1	15	.82
Avoidant	SRAvoid	2A	18	.89
Melancholic	DFMelan	2B	19	.92
Dependent	DADepn	3	14	.81
Histrionic	SPHistr	4A	17	.83
Turbulent	EETurbu	4B	17	.87
Narcissistic	CENarc	5	16	.75
Antisocial	ADAntis	6A	14	.78
Sadistic	ADSadis	6B	14	.80
Compulsive	RCComp	7	13	.67
Negativistic	DRNegat	8A	18	.86
Masochistic	AAMasoc	8B	18	.90
Severe Personality Pathology				
Schizotypal	ESSchizoph	S	21	.89
Borderline	UBCycloph	C	20	.91
Paranoid	MPParaph	P	16	.84
Clinical Syndromes				
Generalized Anxiety	GENanx	A	13	.82
Somatic Symptom	SOMsym	H	10	.84
Bipolar Spectrum	BIPspe	N	13	.71
Persistent Depression	PERdep	D	21	.93
Alcohol Use	ALCuse	B	8	.65
Drug Use	DRGuse	T	11	.83
Posttraumatic Stress	P-Tstr	R	14	.86
Severe Clinical Syndromes				
Schizophrenic Spectrum	SCHspe	SS	21	.86
Major Depression	MAJdep	CC	17	.92
Delusional	DELdis	PP	14	.81

Source: Adapted from *Millon Clinical Multiaxial Inventory–IV Manual*, by T. Millon, S. Grossman, & C. Millon, 2015, Minneapolis, MN: Pearson.

personality patterns as well as clinical symptomatology. In other ways, the MCMI nicely complements the MMPI as the MMPI focuses primarily on what used to be Axis I disorders, whereas the MCMI was designed specifically to assist in describing personality and diagnosing personality disorders. One important advantage of the MCMI is that it is considerably shorter than the MMPI-2 (195 versus 567 items) and even the MMPI-2-RF (388 items) and yet provides a wide range of information. The MCMI takes only 20 to 30 minutes to complete; however, the research base, validity studies, and options for interpretation are clearly more extensive for the MMPI than for the MCMI. Neither instrument should be considered to provide diagnosis. Instead, both provide considerable information relevant to diagnosis. In this sense, they place the clinician in the right diagnostic ballpark, but he or she must then integrate this with other information to make the final diagnosis. In other words, tests (and computer reports) do not diagnose (or make decisions); only practitioners can perform this function.

Factors that greatly assist in useful interpretation are familiarity with the theoretical constructs as well as experience with relevant clinical populations. Theoretical knowledge can be greatly assisted through familiarity with Millon's (2011) *Disorders of Personality* as well as the diagnostic criteria of the fifth edition of the *DSM* (*DSM-5*; American Psychiatric Association [APA], 2013). Although the emphasis on clinical populations implies that the MCMI is intended for psychiatric populations and should not be used with normal persons or those who merely have mild pathology, more attention has been paid in recent editions of the test to the continuum of personality traits from normal to abnormal. Different types of interpretations should be emphasized for persons who score at or above the designated cutoff scores (75 and 85), as compared to those who have mild "elevations" on the scale but who are still clearly below the formal cutoff.

HISTORY AND DEVELOPMENT

Development of the Original MCMI

Shortly after Millon published his 1969 text, *Modern Psychopathology*, fellow professionals urged him to develop an instrument that would operationalize and measure the dimensions of personality as outlined in the book. By 1972, an initial form was developed: the Millon–Illinois Self-Report Inventory (MI-SRI). Over the next five years, the items were further developed, refined, and coordinated with the upcoming personality disorders that were later to be incorporated into the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-III*; APA, 1980). When the initial refinements were completed, the test was published and renamed the Millon Clinical Multiaxial Inventory (MCMI; Millon, 1977).

The formal development of the original MCMI used a combination of rational theory and empirical procedures. The first step was the development of a large pool of face valid questions—a total of 3,500 items derived from Millon's (1969) theories. These were then rationally grouped into 20 different scales. The test developers initially reduced the number of items by rewording those that were poorly worded and removing those that were redundant. Further refinement was done empirically by

having patients rate the clarity and difficulty of the items. A further procedure involved having clinicians regroup the items into scales to evaluate the extent to which these scales related to those originated by the test developers. Based on these procedures, the items were then grouped into two equivalent provisional research forms, with 556 items in each form. The forms were administered to 200 patients, and their responses were evaluated for their endorsement frequency and item-scale intercorrelations. The highest within-scale item intercorrelations were retained, and items that were endorsed either very frequently (>85%) or very rarely (15%) were eliminated. The research form was thereby reduced to a test composed of 289 items.

The 289-item research form was given to 167 clinicians who blindly rated 682 of their patients on 20 different variables after having given them the form. The amount of endorsement frequency and the degree of scale overlap were then used to reduce the items from 289 to 154. Based on this initial validation procedure, three scales were dropped (Sociopathy, Hypochondriasis, and Obsession-Compulsion), and three new scales were developed and added (Drug Abuse, Alcohol, and Hypomania). This brought the total number of surviving items to 175, with 733 different keyings on the 20 different scales.

The scales were initially standardized on 1,591 clinical subjects used in the construction phase of the test. This sample was used to establish the optimal cutoff scores for determining the presence or absence of certain characteristics. A group of 297 non-clinical subjects was used to establish the responses of a normal comparison group. In 1981, the MCMI responses of 43,218 patients were reviewed to further refine and recalculate the cutoff scores.

One feature of the MCMI and its revisions is the use of cutoffs related to Base Rate (BR) scores to designate the presence or absence of a particular characteristic. It is important to note that the MCMI does not use *T* scores. Similar to a *T* score, the BR score is essentially a means of transforming a raw score into a more meaningful score for interpretation. However, BR scores are derived from the percentage of a population that has been deemed to have a certain characteristic or syndrome. For example, 17% of a psychiatric population can be considered to have clear characteristics of a dependent personality whereas only 1% is considered to have clear features of a sadistic personality. This means that decisions regarding client characteristics are made when a client scores in a range that is consistent with either of these two syndromes. However, the relatively more frequent psychiatric disorders with high BRs (i.e., Dependent) require relatively lower cutoff points than those rare disorders with low BRs (i.e., Sadistic). Millon set BR scores of 85 to indicate that the characteristic(s) in question was definitely present; that is, in a patient population given the test, the minimum BR score for those individuals who had that disorder as a primary diagnosis was set to 85. A lower BR score of 75 indicated that some of the features were present; that is, in a patient population given the test, the minimum BR score for those individuals who had that disorder, but not necessarily as their primary diagnosis, was set to 75. Additional cutoff or anchor points were set at 35 to represent the median score for normal or nonpsychiatric groups, and at 60, the median for full, heterogeneous psychiatric populations. This BR approach has been theoretically encouraged by a number of authors (Finn, 1982; Widiger & Kelso, 1983) and empirically demonstrated to increase diagnostic accuracy when compared with the more frequently used *T* score approach (Duthie & Vincent, 1986).

Development of the MCMI-II

The MCMI-II (Millon, 1987) maintained most of the features of the original MCMI. Its development was motivated by a need to incorporate additional research and theory on personality disorders while remaining aligned with the criteria outlined in *DSM-III* and the revised third edition of the *DSM* (*DSM-III-R*; APA, 1987). In addition, between 40 and 50 of the original MCMI items were found to be expendable. Items were developed for two new scales, in part by dividing the previous Negativistic scale into separate scales for Passive-Aggressive (Negativistic) and Self-Defeating. Similarly, the earlier Antisocial-Aggressive scale was divided into an Antisocial scale and an Aggressive/Sadistic scale. Additional items were generated with procedures similar to those used for the original MCMI. This resulted in an MCMI-II Provisional Form of 368 items, which was given to 184 patients who had been carefully diagnosed using *DSM-III-R* criteria. Items were retained or deleted based on the extent to which they could differentiate relevant diagnostic criterion groups. Like the earlier MCMI, the MCMI-II totaled 175 items, but they were keyed on 22 (as opposed to only 20) different scales. In an attempt to reduce scale intercorrelation, individual items were given weightings of 1, 2, or 3 points, based on their relative importance for the specific scales they were being keyed on. Optimal BR cutoff scores were based on a standardization group of 1,292 patients who had a wide variety of presenting problems.

Development of the MCMI-III

Ongoing research, new conceptual developments, and the publication of the *DSM-IV* (APA, 1994) contributed to the MCMI-II's revision into its third version, the MCMI-III (Millon, 1994, 1997; Millon, Millon, Davis, & Grossman, 2006). With procedures similar to those used for the MCMI and MCMI-II, a provisional 325-item test was developed; Depressive and PTSD scales were added. The Self-Defeating and Sadistic Personality Disorder scales were maintained, although these diagnoses were eliminated from the *DSM-IV*. The final MCMI-III still totaled 175 items, but 90 of the items from the MCMI-II were "changed" (85 remained the same). Actually, most of the changed items remained essentially the same in their primary content; the alterations related mostly to increasing the severity of the symptoms. This was done to decrease the number of people endorsing particular items, in the hope that the MCMI-III would be more selective in suggesting pathology. In addition, the items per scale were reduced by half and the number of keyings was reduced from 953 on the MCMI-II to only 440 for the MCMI-III. The possible ratings per item were reduced from 1, 2, or 3 to either 1 or 2. Optimal BR cutoff scores were derived from a standardization sample of 1,079 clinical patients who had come from a diversity of backgrounds and treatment settings. In addition, facets, or subcomponents, were developed for each of the personality scales to allow for a more nuanced interpretation of scale elevations (S. Grossman & del Rio, 2005; Millon et al., 2006).

Development of the MCMI-IV

With the publications of Millon's new edition of *Disorders of Personality* (2011) and of the fifth edition of the *DSM* (*DSM-5*; APA, 2013), as well as changes discovered in

the base rates of certain disorders in the latest norming of the MCMI-III in 2009, a need for another revision emerged. In addition to some minor terminological changes, a new personality scale was added: Exuberant/Turbulent Personality Disorder. More importantly, however, a major shift in the conceptual framework was made that aligned more closely with Millon's theoretical ideas. Because of the focus on clinical samples, many test users previously assumed that the MCMI only explained severe pathological personality functioning. However, the MCMI-IV aims to clarify that clinical samples come with a range of personality functioning (and dysfunction), so that the instrument can be used to describe continua of personality traits, from normal to abnormal (functional to dysfunctional). Although each personality scale was previously conceptualized using the pathological language of personality pathology (e.g., Narcissistic scale, Avoidant scale), now each personality scale has been labeled with a title that represents the continuum from normal to abnormal personality traits. For example, what used to be the Histrionic scale is now the SPHistr scale, representing the continuum from Sociable to Pleasuring to Histrionic. The ADSadis scale represents the continuum from Assertive to Denigrating to Sadistic. For the development of the MCMI-IV, a provisional 284-item test was developed. The final MCMI-IV totals 195 items, with 112 items from the MCMI-III retained and 83 new items added (which replaced 63 of the original 175 MCMI-III items). The new items were developed primarily to expand the scope of the test and make it more contemporary. Optimal BR cutoff scores were derived from a standardization sample of 1,547 clinical patients who had come from a diversity of backgrounds and treatment settings.

Theoretical Considerations

The development of the four versions of the MCMI has been largely guided by Millon's theories of personality. One of his core principles is the intersection of the polarities of pleasure-pain, active-passive, and self-other (R. G. Davis, 1999; Millon, 2011; Millon & Davis, 1996; Millon et al., 2006; Strack, 1999). The pleasure-pain polarity relates to the fundamental evolutionary task of all people in that they must struggle to exist/survive, with the seeking of pleasure being the mode of life enhancement and avoidance of pain the mode of life sustaining. The passive-active polarity relates to the fundamental evolutionary mode by which individuals use various efforts to adapt to their environment or adapt their environment to themselves, with passive representing the modifying of the self to the environment and active representing the modifying of the environment to meet one's needs. The self-other polarity relates to the degree to which individuals invest in other people as well as themselves, ranging from fully self-propagating (self) to fully other-nurturing (other). Each of these polarities can be used to describe differences in personality organization for normal persons as well as those with personality disorders. For example, normal levels of functioning can occur on the active-passive dimension, but when either an active or a passive style is exaggerated, then the person's behavior can become dysfunctional. Thus, schizoid and avoidant personality disorders are extreme in the direction of passivity. In the self-other dimension, dependent and histrionic personalities are highly oriented toward others, whereas the narcissistic personality is extremely self-oriented. Many of the personality styles can be simultaneously portrayed on each of the three polarities. For example, the histrionic style

is quite active and is both other- (dependent) and pleasure-oriented. In some cases, the person is ambivalent on one or more of these dimensions, thereby resembling a person with a passive-aggressive style who is overtly passive and compliant but covertly expresses conflict and anger. Considerably more detail on these polarities, along with other aspects of personality disorders, can be found in Millon's (2011) *Disorders of Personality: Introducing a DSM/ICD Spectrum from Normal to Abnormal*.

Another important point relates to both the test's development and its implications for interpretation: The personality styles are not mutually exclusive. For example, a person with an antisocial style might be frequently uncomfortable with underlying anger and antisocial impulses and thus express them in passive-aggressive modes. This overlap also explains why the diagnosis of personality disorders has been plagued with poor interdiagnostician agreement (poor discriminant validity; R. F. Bornstein, 1998). The expected overlap among characteristics is one reason that the test developers were not overly concerned that many of the scales were highly correlated. Also, the overlap that was present seemed to occur in theoretically consistent patterns. From a practical perspective, this means that combinations of scale elevations can be used to give added meaning to each other. For example, a high score on the Antisocial scale, in combination with an elevation on the Sadistic scale, clearly suggests that the person will act out his or her antisocial feelings in a predictable and potentially dangerous manner. This activity would have very clear implications for case management and treatment planning.

Further, scale elevations should always be placed into the context of the person's life. A high score is not diagnostic of a personality disorder in and of itself. If a person can find an appropriate niche where the expression of his or her personality style is not dysfunctional, that person should not be considered "disordered." Thus, the distinction between a personality "style" and an actual personality "disorder" should be stressed. For example, a salesman with a narcissistic antisocial style might be able to optimize these traits in a way that makes him quite occupationally successful. The diagnostic criteria for personality disorders specifically state that there must be an enduring pattern leading to "clinically significant distress or impairment in social, occupational, or other important areas of functioning" (APA, 2013, p. 646). If there is no or little distress or impairment, a personality disorder should not be diagnosed. This point is particularly crucial for the Compulsive (7), Histrionic (4), and Narcissistic (5) scales, since these scales are often elevated among persons without significant psychopathology (Craig, 2005). Thus, it is crucial for clinicians to determine whether a given personality style suggested by the MCMI has actually led to distress and/or impairment.

Finally, the different categories of scales (Clinical Personality Patterns, Severe Personality Pathology, Clinical Syndrome, Severe Syndrome) are conceptually and clinically related (see Table 9.1). The first two categories relate to personality styles and personality disorder diagnoses but are separated to designate the greater levels of severity for the schizotypal, borderline, and paranoid conditions. As was pointed out, however, any of these personality styles is not a disorder unless there is distress and/or impairment. The second set of categories is intended to measure the type and level of distress and thus relate more to what used to be Axis I diagnoses. These categories represent the results of expression of personality styles that are not working well for the person. For example, if the narcissistic antisocial salesman just mentioned tries to

act toward his wife as he does toward his business contacts, she may file for divorce. His means of coping with this outcome may be abuse of alcohol. In contrast, an individual with a dependent avoidant style who is undergoing a divorce would be more likely to respond with a major depression. This difference underlies the essential interrelationship between personality and expressed symptomatology. It also points out that the MCMI can help to establish the presence of a personality disorder diagnosis by noting the type and degree of distress and impairment as expressed in elevations on the scales in the Clinical Syndrome and Severe Syndrome categories.

RELIABILITY AND VALIDITY

Reliability and validity studies on the MCMI indicate that it is generally a well-constructed psychometric instrument. It should be noted that some of the studies reported here are on the MCMI-III, as the MCMI-IV is such a new measure. For the MCMI-IV, measures of internal consistency have been particularly strong. Alpha coefficients exceeded .80 for 20 of the 25 (nonvalidity) scales, with a high of .93 for the Persistent Depression scale and a low of .65 for Alcohol Use (Millon et al., 2015), as shown in Table 9.1. Test-retest reliabilities have been moderate to high. The MCMI-IV manual reports that over a median 13-day interval, test-retest reliability had a median of .85 (the high was .93 for Histrionic, and the low was .73 for Delusional). For the MCMI-III, Craig (1999) summarized three data sets on test-retest reliabilities ranging from 5 days to 6 months by stating that the median reliability was .78 for the Personality scales and .80 for the Clinical Syndrome scales. Much longer-term MCMI-III test-retest reliabilities spanning 4 years ranged from a high of .73 for Negativistic to a low for Dependent of .59 (Lenzenweger, 1999). This is roughly equivalent to other stable dimensions of personality.

Because the personality scales theoretically represent enduring, ingrained characteristics, they should have greater stability than the clinical scales, which are based on more changeable symptomatic patterns. In some cases, this has been found to be true; in others, little difference has been found. Studies on the MCMI-I indicated the theoretically expected higher stability for the personality scales as opposed to the clinical scales (Piersma, 1986). In contrast, the Craig (1999) summary found very little difference between the personality and clinical scales, despite an extended retesting interval. Similarly, the MCMI-IV manual reports a mean stability of .86 for the personality scales and an only slightly lower mean stability of .82 for the clinical scales. The duration between testings was quite short, though. This finding suggests that the original MCMI may have had the theoretically higher temporal stability for the personality scales than the clinical scales, but later versions have roughly equivalent temporal stabilities between the two categories of scales.

One central issue when evaluating the validity of the MCMI is the extent to which validity studies on previous versions can be generalized to the newer versions. With appropriate caution, some transferability can be justified because the correlations between the MCMI-III and MCMI-IV scales are moderately high. Specifically, the median scale correlation between corresponding MCMI-III and MCMI-IV scales is .87. Only three scales have a correlation below .70, with Narcissistic being the only

one with a significantly questionable relationship at .51. The highest correlation is for Major Depression, at .97. Comparisons for the Turbulent scale could not be made because it was uniquely developed for the MCMI-IV.

More than 20 factor-analytic studies have been performed on the various MCMI versions, and these have generally supported the keying of the items (Retzlaff, Lorr, & Hyer, 1989) as well as the clustering of the factors around Millon's conceptualization of psychopathology (Choca, Retzlaff, Strack, Mouton, & Van Denburg, 1996; Choca & Van Denburg, 2004; McCann, 1991). Factor analysis of the MCMI-II generally supported the organization of the scales. The most extensive published factor analysis involved 769 cases and resulted in an eight-factor solution (Millon, 1987). The largest factor accounted for 31% of the variance; was related to general Maladjustment; and involved depressed affect, impaired interpersonal relationships, low self-esteem, and unusual cognition and self-behavior. The next two largest factors were Acting Out/Self-Indulgent (13% of the variance) and Anxious and Depressed Somatization (8% of the variance). The final factors, listed according to progressively decreasing proportions of the variance, were Compulsively Defended/Delusional Paranoid, Submissive/Aggressive Sadistic, Addictive Disorders, Psychoticism, and Self and Other Conflictual/Erratic Emotionality. Craig and Bivens (1998) performed a factor analysis on the MCMI-III using 444 outpatients and found three factors they labeled General Maladjustment, Paranoid Behavior/Thinking with Detached Acting Out, and Anti-social Acting Out. Rossi, Elklit, and Simonsen (2010) found a four-factor solution the best fit with a large sample. The four factors related to emotional dysregulation, antagonism, extraversion, and impulsivity.

A variety of correlations have been made between the MCMI and various related instruments, including the Beck Depression Inventory, General Behavior Inventory, Michigan Alcoholism Screening Test, State-Trait Anxiety Inventory, Symptom Checklist—90, and MMPI (Millon, 1994, 1997; Millon et al., 2006; Millon et al., 2015). Many are listed for the MCMI-III in its manual (Millon et al., 2006), offering compelling evidence for its convergent validity. Interestingly, for the MCMI-III, Hesse, Guldager, and Holm Linneberg (2012) found that the clinical syndrome scales converged better with a structured interview-based diagnosis when in raw score form than when converted to base rate scores, though these were still adequate. For the MCMI-IV, some of these correlations with other measures are reported in detail in the MCMI-IV manual. Representative findings include expected correlations between Brief Symptom Inventory (BSI) subscales and MCMI-IV scales: Major Depression (CC) and the Depression subscale on the BSI correlated .75; Persistent Depression (D) and the Depression subscale on the BSI correlated .77; and Somatic Symptom (H) and the Somatization subscale on the BSI correlated .58. Similarly, moderate correlations were found between MCMI-IV scales and expected MMPI-2-RF scales. For example, the Paranoid (P) scale was related to RC6 (Ideas of Persecution; .58). The Antisocial (6A) and Sadistic (6B) scales were related to RC4 (Antisocial Behavior; .64 and .47, respectively). For the most part, correlations between the MCMI and external criterion instruments have been in the expected direction.

One of the important and relatively unique contributions of the MCMI has been the development and availability of data on its diagnostic efficiency. This is usually calculated by designating BR scale scores of 75 and/or 85 as *test positives* and

comparing these with clinician ratings of whether the characteristics predicted by the scale scores actually matched these clinician ratings. In some settings, however, it is important to take into account the frequency by which a disorder occurs in that setting (base rate of the disorder). For a test to be effective, it must diagnose a disorder more accurately than the chance occurrence as determined by the BR. For example, forensic and/or substance abuse treatment facilities usually have high numbers of persons with antisocial personality styles. In these cases, calculation of the *positive predictive power* of the MCMI for the particular setting is recommended. Essentially, positive predictive power is a calculation of the probability that a test score accurately indicates the presence of a characteristic or diagnosis based on some other measure such as clinical ratings. Such a calculation involves a formula (Gibertini, Brandenberg, & Retzlaff, 1986; see Millon, 1994, pp. 41–43) in which prevalence rates must be inserted (derived from knowledge regarding a specific client population), along with sensitivity and specificity data (available in the MCMI-IV manual). Such a calculation provides practitioners with an estimate of the extent to which the instrument performs beyond merely base rate levels. For example, if the prevalence (or base rate) of antisocial personalities is .25 but the positive predictive power of the MCMI is .76, the difference (.76 – .25) of .51 indicates that the incremental validity of the instrument is .51 above merely base-rate or chance predictions. This emphasis on levels of certainty, with its implications for actual clinical decision making, is one of the strong features of the MCMI.

Calculations of the positive predictive power of the MCMI-IV indicate good predictive power ranging between about .30 and .68, depending on the scale and the prevalence of the problem in the population (Millon et al., 2015). This finding is supported by R. Rogers et al. (1999), who pooled existing data on the convergent/discriminant validity of the MCMI-II scales and found good support for Avoidant (2A), Schizotypal (S), and Borderline (C); and moderate support for Schizoid (1), Dependent (3), Histrionic (5), Antisocial (6A), Aggressive (6B), Negativistic (8A), Self-Defeating (8B), and Paranoid (P). Little support was found for Compulsive (7%). Positive predictive power for the MCMI-III Axis II scales indicated that the highest accuracy was found for the Dependent (81%), Paranoid (79%), and Compulsive (79%) scales (Millon et al., 2006). In contrast, relatively low positive predictive power was found for the Masochistic (30%), Negativistic (39%), and Depressive (49%) scales. A similar study by R. Davis, Wenger, and Guzman (1997) found that the highest positive predictive power was found for Dependent (.81), Paranoid (.78), and Compulsive (.79), whereas the lowest was found for Masochistic (.30), Negativistic (.39), and Depressive (.49). The low predictive power for Masochistic and Negativistic is consistent with the fact that Masochistic/Self-Defeating was entirely deleted from the *DSM-IV* and Negativistic and Depressive were relegated to the appendix. Deleting these two disorders was thought to be necessary since the *DSM-IV* considered them to be poor diagnoses in the first place; and it also meant that, for the masochistic diagnosis, clinicians did not have the assistance of *DSM-IV* guidelines. Both studies concluded that comparisons of the three generations of the MCMI generally indicate progressive increases in its psychometric characteristics in general and, more specifically, in its diagnostic accuracy (Millon et al., 2006; Millon et al., 2015).

ASSETS AND LIMITATIONS

The strategy of developing the MCMI was innovative and should be commended. The history and development section outlines how this has involved a combination of theoretical-conceptual, internal-structural, and external criterion procedures. Each of the procedures has progressed in a stepwise manner; only those items that survived the previous steps were retained. The result has been an instrument that adheres closely to theory, demonstrates good reliability, and, given the limitations of many of the constructs it is measuring, has shown excellent internal and external validity. The use of BR scores has been a noteworthy innovation and has probably resulted in increases in diagnostic accuracy. However, difficulties have been noted related to the extensive item overlap and low level of interdiagnostician agreement among clinicians using methods such as structured interviews and the MMPI. Additionally, while the scale names have improved with the latest version, they are not as user-friendly as they could be.

As pointed out previously, the MCMI is a relatively time-efficient test that potentially produces a wide range of information. Of central importance, this information focuses not only on clinical symptomatology but also on the more enduring and potentially more problematic personality disorders. These personality disorders can frequently be overlooked. Practitioners might overlook them because (a) the client is more likely to express concern over more overt symptoms, (b) clinical syndrome symptoms are more likely to be foreign to clients, rather than ingrained as part of their enduring pattern of interacting with the world, and (c) personality styles are often more hidden and must be inferred. Clients can feel the emotional pain of symptoms but are rarely aware of the recurring patterns of behaviors and cognitions that are frequently at the core of the development and maintenance of these symptoms. In addition to knowing the client patterns that lead to symptoms, considerable literature supports the usefulness of knowing a client's status related to personality disorders. For example, a personality disorder diagnosis suggests that the client is at risk for interpersonal difficulties; these difficulties may complicate the therapeutic relationship and alter the course of other clinical disorders (R. F. Bornstein, 1998). Turkat (1990) estimated that 50% of clients seeking psychotherapy meet the criteria for a personality disorder. Thus, the MCMI inhabits a crucial niche in objective assessment because it has been designed to better understand personality dysfunction.

Despite the assets of the MCMI, there are a number of inherent difficulties in the assessment of personality disorders. One central issue is that there is no benchmark or gold standard with which to compare the MCMI assessments. Individual clinicians relying on interview information generally have low interdiagnostician agreement (median kappa = .25; J. Perry, 1992). This is sufficiently low enough that it would be unacceptable in any other area of psychological research. Similarly, formal instruments such as the MMPI, MCMI, and structured interviews have shown little agreement (H. Miller, Streiner, & Parkinson, 1992; Streiner & Miller, 1990), which makes it difficult to judge the "true" accuracy of MCMI personality disorder assessment. Several attempts to deal with this have been made. R. F. Bornstein (1998) urged diagnosticians simply to accept that "our ability to describe different personality disorders has outstripped our ability to diagnose them accurately in real-world clinical settings" (p. 334). His solution is to decide whether a client has a personality

disorder (which does have very good interdiagnostician agreement; Loranger et al., 1995) and then rate the intensity and level of impairment of various personality characteristics. For example, it might be decided that a client has a personality disorder with dependent (moderate intensity, low impairment levels) and histrionic (low intensity, moderate impairment) features. In contrast, Westen and Shedler (1999a, 1999b; Shedler & Westen, 2004) pointed out that the actual process of personality disorder diagnosis by clinicians occurs when they infer personality characteristics from client narratives and match the extent to which these inferences match prototypical conceptions of personality disorders. By rank-ordering composite descriptions of prototypical personality disorders, they developed the Shedler-Westen Assessment Procedure—200 (SWAP–200). Clinicians can then rate a particular client and note the extent to which that client meets the ideal descriptions in the SWAP–200. Initial studies have shown the scale has high alpha (above .90 for 14 of 15 diagnoses), good convergent/divergent validity, and supportive factor analysis (Westen & Muderrisoglu, 2006). These alternatives are consistent with a growing consensus that the personality disorders are not categorically dichotomous (present/nonpresent) but rather occur dimensionally on a continuum (Widiger & Trull, 2007). Future research might compare these procedures with MCMI-IV data and especially determine the extent to which the MCMI-IV might demonstrate incremental validity beyond them.

A related issue is that some of the diagnostic criteria incorporated into the MCMI items are closely tied to the *DSM* criteria, whereas others are more closely linked to Millon's theories. In some cases, these criteria are similar; in others, the criteria are different. This inconsistency related to criteria has led to some controversy regarding the relative advantages and disadvantages of having different criteria (Flynn, McCann, & Fairbank, 1995). One disadvantage is that, in many cases, the MCMI should not be considered a *DSM* measure, even though the titles of the scales may lead practitioners to think that it is (Wetzler & Marlowe, 1992; Widiger, Williams, Spitzer, & Francis, 1985). However, the differences between the *DSM-5* and the MCMI-IV may actually be an advantage for some of the scales/disorders because the *DSM* criteria have been criticized as being both insufficiently related to theory and clearly inadequate in some areas. Because the MCMI has not strictly adhered to the *DSM* criteria, it can work to remedy some of the *DSM*'s perceived inadequacies. For example, the *DSM-5* diagnosis for antisocial personality disorder has attained high interrater reliability but has done so by sticking closely to clear behavioral criteria primarily related to overt acts against society. The more intangible but crucial issue of poor conscience development has not been sufficiently addressed, which has led to accusations that the *DSM* criteria relate more to a "criminal" disorder than to a "personality" disorder. The theory behind the MCMI-IV antisocial personality disorder items stresses both overt behaviors and the relative lack of conscience, and this conceptualization is reflected in the item content.

One further issue relevant to the diagnosis of personality disorders is the difficulty in distinguishing state and trait. Theoretically, clinical syndromes relate primarily to states and personality disorders relate to traits. In reality, they are highly interdependent, and it is often difficult to separate them. State (clinical) MCMI elevations seem to be closely related to scores on trait (personality) scales. For example, J. Reich and Noyes (1987) found a 50% decrease in MCMI personality disorder prevalence estimates when the MCMI was given during the recovery phase as opposed to measures given during

the acute phase. Elevations in MCMI-II personality scales were also demonstrated to increase the more state-related MMPI-2 F (and other validity) scales (Grillo, Brown, Hilsabeck, Price, & Lees-Haley, 1994). Given this state/trait distinction, it would be predicted that the trait/personality scales would be more stable than the clinical/state scales; yet, in many cases, this has not been demonstrated to be so. A number of sources, including the MCMI-III and MCMI-IV manuals (Craig, 1999; Millon, 1994, 1997; Millon et al., 2006; Millon et al., 2015), have demonstrated little difference in the test-retest reliabilities between the two categories of scales. Collectively, these observations indicate that state and trait measures are quite interdependent. Due to this finding, the temporal stability of the personality disorders has been questioned (Widiger & Trull, 2007). To account for the variable stability of personality disorders, Millon developed, for some of the MCMI-II and MCMI-III scales, a number of adjustments that work similarly to the *K* correction on the MMPI. Similar to the MMPI *K* correction, though, it is unclear and controversial as to how effectively they achieve their purpose.

Because of the MCMI's reliance on the *DSM*, efforts have been made to incorporate changes that parallel the ongoing developments of the *DSM*. Adjusting to these changes has the advantage of keeping the MCMI current with changing diagnostic criteria, but it has also meant that the MCMI has been revised relatively frequently (Millon, 1977, 1987, 1994; Millon et al., 2006; Millon et al., 2015). In contrast, the MMPI and Personality Assessment Inventory (PAI) have been through far fewer revisions. The relative frequency of changes to the MCMI means that it takes time before sufficient research is available to establish the relation between the new version and the previous one and the implications this may have for interpretation.

An important consideration is whether the MCMI measures actual personality “disorders” or rather personality “style.” As indicated previously, the MCMI-III (and presumably now the MCMI-IV) measures of histrionic, compulsive, and, possibly, narcissistic traits do not seem to be measuring actual disorders but more styles (Craig, 1999). Choca and Van Denburg (2004) preferred to think of the various scales as referring to personality “style” because the inference to disorder requires more information than can realistically be found in scale elevations on any single measure. Persons with certain personality styles may have been able to find an occupational and/or interpersonal niche that allows them to function adequately. For example, a schizoid or avoidant personality may work quite well as a night watchperson. Thus, the inference from style to disorder must be made by the individual practitioner and not by the test. Practitioners who look for the test to include actual diagnosis are overextending its use beyond realistic expectations.

A further issue with the MCMI has been extensive item overlap. The original MCMI (MCMI-I) had its 175 items arranged on 733 different keyings, and the MCMI-II had an even greater 953 keyings. Thus, because many of the items were used to score numerous scales, there were frequently high scale correlations. For example, the MCMI-I's Borderline and Dysthymia scales shared 65% of their items and were highly correlated (.95). Given these characteristics, practitioners might be justifiably concerned that some of the scales were measuring constructs that were too similar and therefore redundant. The defense of the high scale intercorrelation has been that many of the constructs are theoretically and clinically similar, and the similarity would,

therefore, be psychometrically reflected in many high scale correlations. Practitioners would then need to look at the relationships between the different scale elevations as a means of fine-tuning their interpretations. For example, avoidant and schizoid personalities are similar in their passivity and interpersonal distance. Clinical lore suggests that many persons initially believed to have schizoid personality traits appear more similar to those with avoidant personality traits as more information is obtained from them. Given the theory and intent of the MCMI, the two scales measuring these styles would be expected to have similar items. Notwithstanding this defense, the MCMI-III attempted to reduce the item scale overlap (and resulting intercorrelations) by reducing the number of items per scale, providing item weightings depending on their relative importance for a scale, and reducing the number of keyings to 440 (Millon, 1994); the MCMI-IV (Millon et al., 2015) includes 392 keyings on the personality pattern and clinical syndrome scales, even though it has an increased number of items. This reduction in keyings seems to have been successful in that none of the interscale correlations reported in the MCMI-IV manual was above .90 and only three were above .80 (Avoidant—Melancholic $r = .81$; Histrionic—Turbulent $r = .81$; Masochistic—Melancholic $r = .85$), with the rest being below .80. This finding suggests that the MCMI-III scales, compared to the previous versions, are measuring somewhat more independent domains.

When interpreting the MCMI, it is sometimes difficult to know where the interpretive information was derived. That information is based on a combination of theory and empirical relationships determined specifically through validity studies of the MCMI itself. Each of these two interpretive sources has developed over a number of years, during which time four versions of the MCMI have appeared. It is often difficult to know whether the interpretations have been empirically based or theory based and whether they have been derived from validity studies done on previous versions of the MCMI. If done on previous versions, it can be rightfully argued that most of the interpretations can be transferred from these earlier versions because there has been continuity in theory and scale development. This continuity is particularly reflected in the moderate to high correlations between the new and the older scales. However, practitioners must struggle with which of the interpretations have been empirically versus conceptually derived, as well as which are obsolete versus still current. This problem is relevant for the MCMI as well as other similar instruments (e.g., MMPI), and it highlights the importance of clinicians working with the test results and integrating them with additional sources. According to Millon (1992), the quality of the interpretive information is dependent on “the overall validity of the inventory, the adequacy of the theory that provides the logic underlying the separate scales, the skill of the clinician, and the interpreter’s experience with relevant populations” (p. 424).

A criticism related to this issue is that the MCMI overdiagnoses and overpathologizes (P. Flynn et al., 1995). For example, Wetzler (1990) noted that MCMI-related diagnoses of personality disorder were 60% higher than diagnoses based on structured interviews. One of the reasons for at least the potential for overdiagnosis among practitioners is the possibly misleading names of the personality scales, which the latest version of the test has attempted to correct. The pathological names of the scales

created the external appearance of clear *DSM* diagnostic categories when they are probably best conceptualized as styles that may or may not reflect an actual disorder. This does not mean that the MCMI cannot be extremely useful in diagnosing personality disorders, but it should be more accurately perceived as placing the practitioner in the correct domain or coming halfway (or more) toward diagnosis. A further problem is that the MCMI does not perform as well on normal populations. Such persons might have moderate elevations that are still below the BR 75 cutoff, and practitioners might be tempted to interpret these elevations in pathological terms. Again, the renaming of the scales on the MCMI-IV is an attempt to combat this tendency. Unfortunately, the MCMI National Computer System computer interpretations for the MCMI-III tended to both reinforce interpretations of moderate “elevations” and suggest that *DSM* diagnoses can be made based on MCMI scores. A related difficulty is that the scales and their related interpretations tend to emphasize a client’s deficiencies without balancing these out with the client’s strengths. The result is likely to be an overly negative description of a client’s functioning. This occurs despite the fact that many aspects of personality styles might be quite adaptive: for example, the easy sociability of the histrionic style or the adaptability and empathy associated with many persons with depressive styles (see R. F. Bornstein, 1998). In addition to overpathologizing, the MCMI-III was found to perform poorly when assessing persons with psychotic disorders (Craig, 1999).

USE WITH DIVERSE GROUPS

One consideration in interpreting the MCMI is the possible influence of gender, age, and ethnicity. Gender influences have been minimized by using separate norms for scoring the profiles of males and females. The gender differences that have emerged on the MCMI are also consistent with prevalence rate estimates. For example, the greater rate of antisocial personalities among males is reflected in the BR scores, which take this greater prevalence rate into account. Some research has suggested the presence of gender bias in certain items of the MCMI-III, especially on the Narcissistic scale (K. A. Lindsay, Sankis, & Widiger, 2000), though overall scale scores do not show differential functioning by gender in general.

Some differences between white and African American psychiatric patients were found on 9 of the 20 MCMI-II scales. African Americans scored especially higher on Antisocial, Narcissistic, Paranoid, Hypomania, and Drug Abuse scales (Choca, Shanley, Peterson, & Van Denburg, 1990). On the MCMI-III, African American and Caribbean adults scored consistently higher on the Antisocial, Narcissistic, and Delusional Disorder scales (Lloyd, 2009). However, the meaning attributed to these differences is unclear. For example, the greater elevations on these MCMI scales may mean that these scores are accurate representations of the more difficult circumstances many African Americans encounter. Accuracy of MCMI-III elevations is supported in that self-descriptions by African American clients closely correspond to expected elevations on the MCMI-III (Craig & Olson, 2001). However, Urgelles (2014) argued that the normative sample for the MCMI-III did not include enough African

Americans (11.1% of the normative sample) to be representative of the national population. The MCMI-IV continues to struggle with representativeness of the normative sample, with only 8.0% identifying as African American and only 11.1% identifying as Hispanic, still lower than the national percentages.

Finally, there do seem to be some age-related differences on the MCMI: In one sample, older persons scored higher on Dependent but lower on Compulsive and Borderline (Choca, Van Denburg, Bratu, & Meagher, 1995). Interpretations among older persons should take these age-related variables into account.

INTERPRETATION PROCEDURE

Effective interpretation of the MCMI-IV requires considerable sophistication and knowledge related to psychopathology in general and personality disorders in particular. At a minimum, practitioners should be familiar with issues related to personality disorders, along with the *DSM-5* criteria. Ideally, practitioners should also have read Millon's (2011) definitive *Disorders of Personality: Introducing a DSM/ICD Spectrum from Normal to Abnormal*, worked with clients with personality disorders, and administered the MCMI-IV to a number of such clients. Clinicians should also be aware of the previously outlined assets and limitations of the MCMI-IV so that they can most appropriately work with the data. In particular, the MCMI-IV does not provide *DSM-5* diagnoses; its use with clinical and nonclinical populations should be clear; it is not particularly helpful in assessing a person's strengths; and there is a possibility that it might overdiagnose personality disorders and be overinterpreted by clinicians.

The set of procedures outlined in this section is recommended for interpreting the MCMI-IV. The discussion of the various scales and codes represents an integration and summary of current research as well as material included in the MCMI-III manual (Millon, 1994) and interpretive guides developed by Choca and Van Denburg (2004), Craig (2001), and Jankowski (2002) as well as the work of Millon et al. (2015). Much of the work is based on the assumption that interpretive criteria will remain somewhat stable from the MCMI-III to the MCMI-IV. The subsections related to treatment planning have summarized material from Dorr (1999); Goncalves, Woodward, and Millon (1994); Millon and Davis (1996); and Retzlaff and Dunn (2003). Each of the 30 MCMI scales is discussed in relation to interpretation, possible interaction with other scales, and implications for treatment planning.

The formal elaboration and separate listing of 2- and 3-point code types are not discussed for several reasons. First, research on the MCMI does not have the well-developed code-type validity literature found for the MMPI-2. Instead, many of the MCMI code-type descriptions are based on a conceptual integration of the implications of clusters of scale elevations. This means that interpreting patterns of scale elevations is a task that individual practitioners can do themselves by rationally considering the meanings of associated scale elevations. For example, an elevation on Antisocial, combined with a corresponding elevation on Sadistic, would clearly indicate the abusive, combative, and impersonal expression of the person's antisocial tendencies. Second, given that there are fully 30 MCMI-IV scales, the total number of possible code types is both unwieldy to list and unrealistic to fully research. However, a short subsection ("Frequent Code Types") under most of the scale descriptions does

briefly describe the meanings attached to some of the more important associated scale elevations. Readers are encouraged to read these descriptions and to expand on their meanings by reading the longer interpretive descriptions for the entire associated scale.

1. Determine Profile Validity

Before interpreting the personality and clinical scales, practitioners must be assured that the client has not over- or underreported symptoms or responded in a random manner. The profile validity can be assessed by noting the pattern of scores on the Modifying Indices (validity indicators):

Random responding is suggested by scores of one or more on the three items of the MCMI-IV Invalidity scale (“True” on items 49, 98, and 160). In addition, the MCMI-IV has added a new scale (W: Inconsistency), which attempts to detect random responding by evaluating 25 pairs of items that are statistically and semantically similar. While it is unlikely that these items be endorsed in an opposite direction, it is semantically possible (e.g., it is possible that “I’m a loner and I don’t mind it,” and “When I have a choice, I prefer to do things alone” could be validly answered in opposite directions). However, higher numbers of these semantically mismatched responses becomes less plausible. Opposite endorsement of an item pair is scored as a single raw score point on the W scale, with scores of 0–8 being acceptable (achieved by about 98% of the normative sample), 9–19 being questionable (achieved by about 2% of the normative sample), and 20–25 being considered invalid.

Underreporting of difficulties on the MCMI-IV is suggested by low scores (raw score less than 34) on Disclosure (X) and Debasement (Z), together with a high score (BR over 75) on Desirability (Y). However, it is sometimes difficult to differentiate persons who are faking good (underreporting) from those who actually have the positive qualities of being cooperative, self-confident, and conscientious. The client’s history is often the best tool for making this distinction.

Fake bad profiles are suggested by a high score (raw score above 178) on Disclosure (X) and a high score (BR above 75) on Debasement (Z). With moderate elevations, this might be a “cry for help” or a client with truly significant psychopathology, but with progressively higher scores (BR above 85), the likelihood of an invalid profile is increased.

It should be noted that BR adjustments for certain scales have been made in an effort to increase MCMI-IV profile validity. The adjustments serve as correction scores in much the same way as the *K* correction serves for the MMPI-2. These adjustments are part of the standard scoring and involve adjustments to several scales based on the levels of Disclosure (if either at the top or bottom 10th percentile) and a value derived from the Generalized Anxiety and Major Depression scales. Extremes on the Disclosure scale can raise or lower the BR scores on all of the personality and syndrome scales (a maximum of 10 BR points up or down on the personality scales and of 5 BR points up or down on the syndrome scales). High values on anxiety and depression (BR > 75) lower the BR scores on the Melancholic (2B, DFMelan), Masochistic (8B, AAMasoc), and Borderline (C, UBCycloph) scales a maximum of 10 BR points. High values on anxiety and depression (BR > 75) lower the BR scores on the Avoidant (2A, SRAvoid) and Schizotypal (S, ESSchizoph) scales a maximum of 5 BR points.

2. Interpret the Personality Disorder Scales

Retzlaff (1995) recommended that, when interpreting the Personality Disorder scales, practitioners should first check to see whether any of the Severe Personality Disorder scales are elevated. If so, this strongly suggests that one or more of the Clinical Personality Pattern scales will also be elevated. However, the high scale(s) on the Severe Personality Disorder section should take precedence over equivalently elevated scales on the Clinical Personality Pattern scales. The Clinical Personality Pattern scales then serve to color or elaborate on the elevation(s) on the Severe Personality Disorder scale(s). The primary focus for diagnosis, then, should be to rely on the Severe Personality Disorder elevation unless elevations on other categories of scales were extremely elevated compared to the Severe Personality Disorder scales. When that occurs, the extremely elevated scales would take on greater interpretive meaning compared to the more moderately elevated Severe Personality Disorder scale(s). (Interpretive descriptions of each of these scales, which are based heavily on the work of Millon, 2011, can be found in the next section.) If there are no elevations on the Severe Personality Disorder scales, practitioners should then interpret any elevations on the Clinical Personality Pattern scales.

The interpretive sections under the Personality Disorder scales are divided into general interpretive descriptions (including the Grossman facet scales and three levels of severity for each scale), frequent code types, and treatment implications. Evaluators need to determine whether the descriptions apply specifically to the individual client, based on how high the scale is elevated, implications of associated scale elevations, and additional data available on the client. Although score ranges are somewhat flexible, in general, marginal elevation (BR around 75–85) represents the middle descriptor for each scale, while extreme elevation (BR above 85) represents the final and most pathological descriptor. Elevations that are below a BR of 75 but are still a primary elevation for the individual reflect the first, least pathological descriptor. The general rule, then, is: The higher the elevation, the more likely that the more pathological descriptors are appropriate. Another consideration is the height of elevated scales relative to other elevated scales. If they are approximately the same height, they should be given equal interpretive weight. If there are 20 or more BR points between scales, however, the lower scale's influence is likely to be subtle, such that it should be used to nuance the meaning of the higher scale.

To help with specific interpretation of the personality pattern scales, the Grossman facet scales are conceptual breakdowns that delineate different aspects of personality for each scale, including cognitive components, interpersonal components, self-image components, and behavioral aspects, among others. These facet scales help specify what aspects of personality styles are present in a client. When interpreting an elevated personality scale, the three Grossman facet scales associated with that overall personality scale should be consulted to add specificity and nuance to the interpretation. For example, a client may have an elevated Antisocial scale and, upon review of the three facet scales, may have elevation on the Irresponsible Interpersonal Conduct (behavior), but not on the Autonomous Self-Image or Acting-Out Intrapsychic Dynamics facet scales. This kind of style may suggest an individual who acts irresponsibly and perhaps manipulatively, resentful about certain things, but not recklessly impulsive or

unconfined by loyalties and attachments. This profile might be found in a very high-ranking executive in a company, who has ruthlessly and successfully worked his or her way to the top. Alternatively, an individual with elevations on all three facets of the Antisocial personality scale would much more likely have impairment in his or her functioning, perhaps also elevated on drug or alcohol use or having committed significant crimes. Thus, the same elevation on Antisocial may have different specific components that help clarify personality functioning.

3. Interpret Clinical Syndrome Scales

Similar to step 2, Retzlaff (1995) recommends that precedence be given to interpreting any elevations on the Severe Clinical Syndrome scales. Sometimes all or most of these scales are elevated, which should not be considered contradictory; rather, these elevations can be used to complement one another. Any elevations on the Severe Clinical Syndrome scales are usually accompanied by complementary elevations on the Basic Clinical Syndrome scales, as well as often the Personality Disorder scales. For example, an elevation on the Severe Clinical Syndrome scale of Major Depression might also have corresponding elevations on Drug Use, Generalized Anxiety, and Avoidant. Interpretations would center on depression but would also include fear of interpersonal involvement, anxiety, and the distinct likelihood that the person is using alcohol as a means of coping with these difficulties. Another example might be a person with an elevation on Anxiety but with a corresponding elevation on Avoidant and Dependent, which suggests he or she is experiencing the anxiety because of conflict between wanting to be accepted and cared for by others and being terrified of criticism and humiliation (often a dynamic in borderline personality disorder). In contrast, another person with an elevation on Anxiety but with a corresponding elevation on Narcissistic is most likely experiencing anxiety because of significant challenges to his or her self-inflated sense of importance and superiority, which may mask a fragile sense of self-worth. This careful interplay between the scales is crucial for accurate and effective profile interpretation.

One of the unique features of the MCMI-IV is that it is an objective test that measures personality styles/patterns relevant to personality disorders. The sections describing each of these scales include subsections on frequent code types (including possible relations with Clinical Syndrome scales) and treatment implications. In contrast, the Clinical Syndrome scale descriptions include descriptions of only the scales, without material on frequent code types or treatment implications, partly because relevant relations with the personality scales are already addressed. In addition, there is already a well-developed clinical literature (and extensive time is spent in most training programs) on treating these clinical syndromes (anxiety, depression, etc.).

4. Review Noteworthy Responses (Critical Items)

The MCMI-IV manual (Millon et al., 2015) has series of Noteworthy Responses, which, similar to the MMPI-2/MMPI-2-RF critical items, are not so much formal scales as they are rationally categorized items that may be important for a clinician to highlight, whether for potential harm/liability reasons or for the potential to

inform differential diagnosis. These are organized around the topics of Traumatic Brain Injury, Vengefully Prone, Self-Destructive Potential, Emotional Dyscontrol, Interpersonally Alienated, Health Preoccupied, Explosively Angry, Eating Disorder, Childhood Abuse, Prescription Drug Use, Self-Injurious Behavior/Tendency, Adult ADHD, and Autism Spectrum. These Noteworthy Responses can be used to organize a semistructured interview around relevant responses. They can also be selectively inserted into a psychological report to provide a more concrete qualitative portrayal of the client's attitudes, affect, and behavior.

5. Provide Diagnostic Impressions

Given the interpretive descriptions of a client's profile (steps 2, 3, and 4), along with any other relevant information, clinicians can formulate the most appropriate diagnosis. Again, it should be noted that the MCMI-IV cannot alone provide a diagnosis. Rather, it can describe styles and clusters of symptoms that may impair functioning and, if they do, point the practitioner in the direction of accurate diagnosis.

6. Elaborate on Treatment Implications and Recommendations

The symptoms reported and reflected in elevations on the Clinical Syndrome scales (Anxiety, Depression, Substance Abuse, etc.) are those that are generally the most salient to the client and thus should be targeted as high priorities. However, these also need to be understood in the context of the client's personality patterns and pathologies. Under each of the personality scales, there are sections with relevant suggestions for treatment recommendations. These can be considered, along with other information, to expand on what would be the most appropriate interventions. Additional useful resources in this process are Chapter 14 of this book ("Treatment Planning and Clinical Decision Making"), Millon and Davis's (1996) *Disorders of Personality: DSM-IV and Beyond*, Millon's (1999) *Personality Guided Therapy*, and Magnavita's (2008) chapter "Using the MCMI in General Treatment Planning."

VALIDITY SCALES

The MCMI Modifying Indices are adequate at detecting random responding, fake bad, and fake good profiles. However, dated research showed that the detection rate appears lower than for the MMPI (Bagby, Gillis, Toner, & Goldberg, 1991), and, as with the MMPI, fake bad profiles are more accurately detected than fake good (defensive) profiles (Fals-Stewart, 1995; Millon, 1987). Using the decision rules for fake bad profiles, the rate of accurate detection runs between 48% and 92% (Bagby, Gillis, Toner, et al., 1991; Retzlaff, Sheehan, & Fiel, 1991; Schoenberg, Dorr, & Morgan, 2003). However, for clients with severe disturbances, high scores on fake bad indices may be more indicative of high distress and a cry for help than an invalid profile (Wetzler & Marlowe, 1990). In contrast to the generally good detection rate for fake bad profiles, persons faking good (defensively) are likely to be detected only approximately 50% of the time (Retzlaff et al., 1991), and clients underreporting their substance abuse seem

particularly good at avoiding detection (Fals-Stewart, 1995). Thus, the MCMI should be used with extreme caution in situations in which individuals might be likely to under-report their psychopathology.

The most useful tool in making these decisions related to validity is a careful consideration of the client's past and current level of functioning. Specifically, a person who may look as though he or she is faking bad, but whose history reveals someone who is dysfunctional, may be merely expressing distress. In contrast, a relatively highly functioning person with the same scores on the Modifying Indexes is much more likely to be faking bad. Conversely, a person with a potentially fake good profile but who also has a high level of functioning may be merely expressing actual confidence, assertiveness, and high self-esteem. A person with a similar profile but a history of interpersonal, legal, and/or psychiatric history, however, is much more likely to be underreporting psychopathology.

Invalidity Index (Scale V)

The MCMI-IV Invalidity Index is composed of three items (numbers 49, 98, 160) that, if endorsed as true, indicate absurd responses. As a result, endorsement of these items strongly suggests that a person has responded randomly. The manual states that one true response should be interpreted as indicating a profile of "questionable validity" and two or more endorsements can clearly be interpreted as an invalid profile. Presumably, the "questionable validity" option is given to suggest that the profile *may* still be valid in the event that the client has misread or randomly responded to only a few items (including one of the three on the Validity Index). Misreading only a few of the items allows for the possibility that most of the items were still responded to accurately. In contrast, Bagby, Gillis, and Rogers (1991) recommended that even one endorsed item be used to indicate an invalid profile. One caution: If a person did respond randomly, there is still a 50% chance that he or she may have gotten "lucky" and none of the three items was answered in a true direction, in which case detection on the Validity Index would be avoided (Charter & Lopez, 2002). In addition, a person wishing consciously to fake responses would be able to notice the absurdity of answering "true" to any of the Invalidity Index questions and would answer them in such a way as to not endorse scorable responses on the Validity Index.

Inconsistency Index (Scale W)

The MCMI-IV Inconsistency Index is composed of 25 pairs of items that are both statistically and semantically similar (though not identical). It is generally expected that most of these will be answered in the same direction, though because they are not identical, it is theoretically semantically possible that an individual would answer several of the pairs in opposite directions. For example, the paired items "I'm a loner and I don't mind it" and "When I have a choice, I prefer to do things alone" are quite similar in their meaning. It is expected that most people would answer both in the same direction. However, if an individual answers them in opposite directions, it may not necessarily be an indication of random or inconsistent responding. It may be an actual, real difference. As a result, there is some room for maintaining a valid protocol with

some level of discrepancy on these paired items. Each pair that is answered in opposite directions is scored as one raw score point on the W scale. The manual states that scores of 0 to 8 (which about 98% of the normative sample achieved) are acceptable, 9 to 19 (which was achieved by about 2% of the normative sample) has questionable validity, and 20 to 25 should be considered invalid.

Disclosure Index (X)

The Disclosure Index was designed to measure whether a client's responses were open and revealing as opposed to defensive and secretive. If the MCMI-IV score on the Disclosure Index is below 21, it most likely indicates a defensive underreporting of psychopathology, with scores below 7 being considered invalid. Low scores may also mean that the person did not read or understand the questions correctly. A further interpretation is that the client is hesitant, reserved, and overconcerned with seeking social approval. However, low Disclosure Index scores on the MCMI-II were not found to be particularly sensitive because people requested to "fake good" still produced generally acceptable Disclosure Index scores (Retzlaff et al., 1991). Thus, when clients do fake good extensively enough to produce a clearly low Disclosure Index, the profile can be considered invalid with a fair degree of certainty.

MCMI-IV Disclosure scores above 60 indicate that the individual has possibly exaggerated his or her symptoms, with scores above 114 being considered invalid. Very high scores would even exceed fairly disturbed psychiatric populations and, therefore, suggest an overreporting (faking bad) of symptoms. Some caution should be taken in that the scale has been found to be fairly tolerant to overreporting (C. D. Morgan, Schoenberg, Dorr, & Burke, 2002). Thus, many persons exaggerating symptoms may still score below the score cutoff of 114. If they do score above this cutoff, it can be fairly safely inferred that they are indeed overreporting their symptoms. The index has been found to function somewhat similarly to the Debasement Index (Blais, Benedict, & Norman, 1995), and so these two indexes may not provide independent information.

Desirability Index (Y)

Similar to the Disclosure Index, the Desirability Index is a measure of defensive responding. Scores above BR 75 indicate the individual has presented in a manner that is unusually moral, interpersonally attractive, extremely emotionally stable, highly gregarious, organized, and with a high respect for the rules of society. Progressively higher scores suggest that the person is concealing crucial details regarding psychological or interpersonal difficulties. The Desirability Index has been shown to function independently from both the Debasement and Disclosure Indexes (Blais, Benedict, & Norman, 1995).

Debasement Index (Z)

As the title of the scale suggests, the Debasement Index reflects the extent to which a person is describing him- or herself in negative, pathological terms. Elevated scores on the Debasement Index might include feelings of being empty or angry, crying

easily, having low self-esteem, possibly being self-destructive, and frequently feeling tense, guilty, and depressed. Thus, the Debasement Index measures characteristics opposite from those on the Desirability Index. They are rarely both elevated, although, on occasion, someone who is unusually self-disclosing may have high scores on both. Scores above BR 85 indicate either a cry for help resulting from acute psychological distress or a fake bad profile. The index has been found to function somewhat similarly to the Disclosure Index (Blais, Benedict, & Norman, 1995).

CLINICAL PERSONALITY PATTERNS

Scale 1: Apathetic-Asocial-Schizoid (AASchd)

The core characteristics measured by this scale relate to the ability to form social relationships, respond to typical social reinforcements, feel in deep and varied ways, and enjoy pleasure and positive emotions. Millon (2011) characterized individuals with elevations on this scale as largely unmoved by emotions, unmotivated to seek out pleasure and relatively unmoved by negative emotions. They are disengaged socially and can be passive interpersonally, though they are generally quite content not to engage in activities, relationships, or other life experiences.

The *Unengaged Interpersonal Conduct* facet scale reflects the degree to which these individuals are remote, aloof, and solitary, as well as unmoved by interactions with others. Elevations on this facet scale reflect the tendency to prefer solitary activities or to blend into the background of social activities. Additionally, these individuals often lack interpersonal sensitivity and the ability to understand nuanced aspects of others' motivations or emotions. Feedback to such individuals may revolve around their preference for being on their own and their comfort doing things that do not involve other people.

The *Meager Intrapsychic Content* facet scale reflects the general vagueness in the experience of interpersonal interactions, both historically and presently. That is, elevation on this facet scale reveals that an individual does not have strong, clear, differentiated experiences with other people on which to base current interactions, and his or her interactions or relations with other people do not leave a lasting, clear impression on him or her. Feedback to such individuals may center on the fact that they are not too affected by other people in a lasting way, they do not tend to hold grudges, and they rarely let other people bother them for an extended period of time.

The *Apathetic Mood/Temperament* facet scale reflects the degree to which an individual is able to experience emotions, including joy, sadness, and anger. The range of emotional experience for those with elevations on this facet scale is extremely narrow, and they rarely if ever become emotionally excited. Feedback to such individuals may focus on their limited need for affection; their stable, steady mood; and their lack of strong emotions.

Apathetic

Individuals who have one of their highest three elevations on the AASchd but not significant elevation (BR < 75) have traits that are considered apathetic. These individuals are self-sufficient and do not need much interaction with or validation from others.

They may engage in significant relationships, though at their core they are likely most comfortable and satisfied when they are on their own. They are often interpersonally reserved and prefer to be alone with their thoughts. Feedback to these individuals may focus on their need for time alone and preference for independent activities.

Asocial

Individuals who have one of their highest three elevations on AASchd with a moderate elevation have traits that are considered asocial. The level of social isolation for these individuals is significantly greater than those who are apathetic, and increasingly these individuals have difficulty understanding the nuances of both their own internal world and interpersonal interactions. Higher elevations relate to the beginnings of odd thinking, as the interaction with others with conventional thinking is quite limited. These individuals are often quite slow in their social interactions, not having many of the automatic reactions that individuals with more social interaction generally have. Feedback to these individuals may focus on their distinct lack of interest in socializing with others, including a very limited need for sexual or romantic encounters. Additionally, feedback may focus on some thinking that is not conventional or is preoccupied on certain topics.

Schizoid

The core characteristic of persons with high elevations on this scale is little or no interest in other people. They spend their lives as loners. They are detached, impersonal, withdrawn, unsociable, reclusive, passive, distant, and have few, if any, friends. They rarely initiate conversation, are indifferent to other people, and rarely seek involvement with others. In family, work, or social situations, they prefer to have a peripheral role. As a result, they frequently function on the margins of society. They have little drive to have their needs met, experience few if any romantic attachments, express little warmth, and are often asexual. Rarely do they experience very much depth of feeling (pleasure, sadness, anger). They are largely indifferent to praise or criticism from others. Their interpersonal distance is not based on a defense stemming from fear of rejection, but is rather their natural and most comfortable way of functioning. They also lack vitality and can be unanimated and almost robotlike in their movements. When they do communicate with others, it is in a vague, distant, unfocused, and often slow manner. Often the direction of their conversation loses its focus, and whatever information is conveyed is delivered in a circuitous manner. As a result, others are likely to see them as strange or “spacy.” They have little self-awareness or insight into the implications of interpersonal relationships. If they are involved in a committed or intimate relationship, a frequent spousal complaint is that there is insufficient closeness, sharing, and understanding.

An asset of this personality style is that these persons typically do not become particularly disturbed by anything. Although they are not particularly involved with or interested in others, when they do interact, they are typically quite comfortable. They are also quite self-sufficient—they are comfortable with spending extensive periods of time alone and may have a rich fantasy life. Their hobbies typically involve activities that require only minimal contact with other people.

Frequent Code Types

Clinical scales that are most likely to be elevated along with AASchd/Schizoid are Anxiety and Schizophrenic Spectrum. This pattern of elevation reflects the sometimes obsessive thinking of individuals with elevations, along with the possibility that brief psychotic states might occur. Personality scales that are often elevated along with AASchd/Schizoid are SRAvoid/Avoidant and ESSchizoph/Schizotypal, with lower frequency of DRNegat/Negativistic, DADepn/Dependent, and RCComp/Compulsive. Each of these scales adds new variations onto the previous description. A corresponding elevation on Avoidant suggests that these persons are not only uninterested in and unskilled at interpersonal relationships but also uncomfortable around others and fear rejection. However, behind their detachment may be a real desire to become involved. If Schizoid and Avoidant are both elevated, the possibility of problem use of alcohol should be investigated (check Alcohol Use and Drug Use scales). Elevations on the Negativistic scale (along with Avoidant) underscore conflicted feelings and possible resentment toward the few interpersonal relationships they have. This resentment centers on a wish that someone would nurture and guide them (especially if Dependent is also elevated), along with fear that they might be rejected. This conflict results in their frequently being moody and nervous. When Compulsive is elevated, these persons are disciplined, well organized, emotionally controlled, meticulous, dependable, and persistent. These characteristics are present in part because these individuals are not significantly affected by emotions and as a result remain disciplined, self-restrained, and proper. They are typically overly polite and even ingratiating toward authority figures but, in contrast, may be somewhat disdainful toward subordinates.

Treatment Implications

The two major goals when working with persons with AASchd/Schizoid elevations are to (1) encourage at least some increase in social interaction and (2) help them to enhance their ability to experience pleasure. However, these goals are difficult to achieve in a client who is neither likely to become particularly involved in the therapeutic relationship nor ready to place much value in exploration and insight. As a result, the prognosis is poor. In addition, many therapists are likely to feel that individuals with schizoid features are not particularly rewarding to work with. Therapists must be prepared for long silences and a distant relationship. Yet any relationship that does develop can be extremely important for clients. Problem solving should be directed at concrete, practical matters. Useful techniques might be audio or video recorded feedback about their behavior, cognitive monitoring, and reorientation of their internal processes. Operant conditioning, however, might prove difficult because they have little capacity for external rewards. Similarly, insight might be unproductive because they are not particularly psychologically minded.

Scale 2A: Shy-Reticent-Avoidant (SRAvoid)

The core characteristics measured by this scale are comfort and fear related to being around other people, feeling judged by others, and “performing” socially. The scale measures how sensitive a person is to feeling evaluated and potentially rejected by

other people and, as a result, patterns of social behavior, from outgoing to inhibited. Additionally, a component of the SRAvoid/Avoidant scale is a sense of how adequate or inadequate the individual feels.

The *Aversive Interpersonal Conduct* facet scale reflects what is generally seen as the hallmark and most identifiable feature of the shy-reticent-avoidant spectrum, the degree to which an individual shies away from social situations or needs to ensure that he or she will be accepted before engaging in them. Individuals who score high on this facet scale distance themselves from situations that involve personal interactions, both on a social level (e.g., parties with a great number of people) and on an individual level (e.g., intimate relationships). They distance themselves from other people in order to protect themselves. Feedback to such individuals may focus on how it can make them quite nervous to interact with other people, especially when those people may reject them in some way.

The *Alienated Self-Image* facet scale reflects the degree to which individuals feel secure and capable versus insecure and inept in social situations. Individuals who score high on this facet scale feel that they are justified in isolating themselves because they are awkward, unappealing, and socially inadequate. Feedback to such individuals may focus on their feelings that they are inferior to others in their social abilities, and that they are often sad and anxious. Additionally, feedback may include a reflection that they likely perceive others as harsh and critical of them.

The *Vexatious Intrapsychic Content* facet scale reflects the internal world of positive and negative interpersonal memories that shape the way individuals interact with others in the world. The internal world of individuals who score high on this facet scale is overwhelmingly filled with negative interpersonal experiences and memories, with very few positive or rewarding interpersonal relationships. They are plagued by negative memories, and they in turn expect very few positive or rewarding interactions with other people. Additionally, unlike some individuals high on AASchd/Schizoid, they do not find peace or comfort in their own internal world. That is, turning away from their external environment protects them from what they expect to be negative interactions, but their internal world is no more pleasant. Feedback to such individuals may focus on their low feelings of self-worth and self-respect, as well as their expectations that interacting with others will not ultimately be a positive experience in any way.

Shy

Individuals who have one of their highest three elevations on the SRAvoid but not significant elevation ($BR < 75$) have traits that are considered shy. These individuals are sensitive to the evaluations of others and always aware of potential for being rejected; as a result, they are hesitant in social and interpersonal situations. They most often have low self-esteem and feelings of social inadequacy, though when comfortable can socialize or connect with other people easily. Feedback to these individuals may focus on their hesitation or unsureness in interacting with others, not knowing if they will be accepted. Additionally, their self-consciousness is likely to be quite evident to them, so feedback may focus on that and their general social unease.

Reticent

Individuals who have one of their highest three elevations on SRAvoid with a moderate elevation have traits that are considered reticent. The level of disengagement from social and interpersonal interactions, as well as a general lack of emotional vitality, is significantly greater than those who are shy, and increasingly these individuals are sensitive to potential rejection and may even become restlessly and irritably phobic of interaction with others. Higher elevations relate to the presence of intense and highly fluctuating mood states, as well as active self-deprecation. They often appear apathetic on the surface, in order to distance themselves from others and from their own emotional sensitivity. Feedback to these individuals may focus on their significant disengagement from other people, as well as their low self-esteem and fluctuating, though often depressed, mood.

Avoidant

Individuals with extreme elevations on both the AASchd/Schizoid and the SRAvoid/Avoidant scales live solitary, often isolated lives. However, those with schizoid elevations are indifferent to relationships, whereas those with avoidant elevations desperately want to become accepted and involved with other people—a desire that is blocked by an intense fear of being rejected and humiliated. They warily scan their environment for threats and continually try to present themselves in as favorable a manner as possible. This is rarely successful; they feel a continual sense of unease, disquiet, anxiety, and overreaction to minor events. Thus, they are frequently preoccupied with intrusive, fearful, and disruptive thoughts. They perceive themselves as socially inept, inferior, and inadequate, and they continually undervalue their achievements. In addition to fear and self-criticism, they frequently feel alone, empty, and isolated. To protect themselves from these fears, they restrict their social environments, constantly maintaining their distance and privacy. This is unfortunate because it undercuts future opportunities for enhancing relationships and places them in a solitary world where they are more likely to reactivate memories of past social rejections. In addition, they rely extensively on fantasy gratification of their needs for affection and anger. Given these dynamics, they are quite likely to fulfill the formal criteria for a social phobia and are frequently depressed. Often they are described as withdrawn, insecure, edgy, fretful, insecure, isolated, and rejected.

The positive side of avoidant elevations is that these individuals can be extremely sensitive to the needs and perspectives of others. They can potentially show considerable compassion and understanding and can be emotionally responsive.

Frequent Code Types

Those who score high on SRAvoid/Avoidant are likely to experience a wide variety of clinical syndromes and disorders. As a result, it is quite common to see elevations in several of the clinical syndrome scales. Among the most frequent associated disorders are generalized anxiety, phobias, and social phobias (check Generalized Anxiety). Depression (check Persistent Depression and Major Depression),

hypochondriacal syndromes, and conversion disorders can also occur (check Somatic Symptom). Personality Pattern scales that can be elevated are Dependent, Schizoid, Melancholic, and Paranoid. A corresponding elevation on Dependent augments the core dynamics of the avoidant in that the person has even stronger needs not only to become involved with others but also to be supported by and given guidance by them. This dual elevation on Avoidant and Dependent should be examined for the possibility of meeting criteria for borderline personality disorder, as it signals a core ambivalent, conflicted dynamic. Avoidant in combination with Schizoid adds the dimension of having a lack of awareness or even of interest in personal feelings. These persons are also likely to be detached, aloof, and apathetic, and they rarely develop strong emotional ties with others. They might have some acquaintances, but they are not likely to have any intimate friendships. Elevations on Negativistic suggest moodiness and resentment combined with significant difficulty in trusting others. These individuals might vacillate between being friendly and cooperative and then being hostile, which might be followed by apologies. Because they would feel uncomfortable with their anger, they might resort to covert expressions of hostility, such as passive obstructionism. Whereas many persons with avoidant characteristics have low self-esteem, those with elevations on both Avoidant and Narcissistic have an inflated sense of importance, overestimating their own value and often resenting others they feel do not recognize their specialness. They are unappreciative of others and justify this attitude by perceiving themselves as special. Situations are framed in such a way as to enhance their own self-worth, and they describe themselves as intelligent, sophisticated, outgoing, and charming. However, their underlying style is avoidant, so their sense of self-importance is defensive in nature, extremely flimsy and easily deflated. Elevations on Antisocial introduce to avoidant individuals' personality a competitive edge that might be expressed in hostile and exploitive behaviors. They would justify these behaviors by fears that others are trying to take advantage of them. They usually describe themselves as self-reliant, strong, realistic, and assertive, and they exhibit a contemptuous attitude toward persons who do not have these qualities. In addition, they are likely to be impulsive, argumentative, guarded, reserved, intimidating, cold, and insensitive to the feelings of others.

Treatment Implications

Avoidant individuals are among the most frequent clients in therapy. A potentially difficult issue is that they reveal only the information they believe will not lead to rejection by the therapist. The central treatment task is to change these clients' self-image, but doing so involves working with interpersonal behavior and helping regulate their mood. Particularly useful techniques would be in vivo exposure based on a graded hierarchy, anxiety management training, cognitive restructuring to challenge thinking errors, assertiveness training, and, possibly, psychopharmacological interventions to deal with anxiety states and possible panic attacks. However, the most difficult challenge is to keep them in therapy long enough to achieve therapeutic gain. Motivating them to remain in therapy requires carefully balancing support, empathy, and trust building while still encouraging them to experience situations that challenge them to work on new behaviors and perceptions. Because their high level of arousal would

likely be the primary reason for their terminating prematurely, techniques of arousal reduction—emotional support, reassurance, relaxation, hypnosis, thought stopping, and supportive interpretation—would be particularly important to use. Typically, these clients make significant therapeutic gains. One area to investigate is the possibility that they are using alcohol or drugs to medicate their anxiety. However, referral to a 12-step or peer support program might be difficult, given their avoidant style; therefore, other forms of intervention should be considered.

Scale 2B: Dejected-Forlorn-Melancholic (DFMelan)

The core characteristics measured by this scale are related to the ability and tendency to experience joy and pleasure versus pain, disappointment, and despair. Additionally, this scale reflects an orientation toward the future that is optimistic versus pessimistic, having hope regarding the future versus an outlook that is filled with despair, hopelessness, and lack of pleasure.

The *Fatalistic Cognitive Style* facet scale focuses on the tendency for individuals to attribute situations and the world in general with an optimistic versus pessimistic lens. Individuals who score high on this facet scale view the world in extremely negative, bleak terms, weighted down by pessimism and a lack of hope that things will improve. They believe that negative things in their lives are irreversible and global, representations of the fact that they are inadequate. Rumination about helplessness and hopelessness are common. Feedback to such individuals may focus on how tough and exhausting it is to view the world in such negative terms.

The *Worthless Self-Image* facet scale focuses on the degree to which individuals feel they are inadequate in general and undeserving of accomplishments or praise. Individuals who score high on this facet scale not only feel worthless and inadequate, but they tend to feel guilty about the fact that they do not possess more positive qualities. When things go well, they dismiss the situation as an anomaly, a fluke, or the result of outside forces, maintaining their deep sense of inadequacy. When things do not go well, this reinforces their general feeling of worthlessness. Feedback to such individuals may focus on their feelings that they are not good enough and do not have the positive qualities that many others do.

The *Woeful Mood/Temperament* facet scale focuses on the degree to which individuals experience dysphoric and gloomy feelings. Individuals who score high on this facet scale are gloomy, unmotivated, and generally morose. They feel empty and sad and may feel particularly guilty, tortured, or irritable. Their sullen demeanor is quite evident in the plodding and reluctant nature by which they engage in life's tasks. Feedback to such individuals may focus on their generally sad and gloomy state, as well as the fact that they have very little joy in their lives.

Dejected

Individuals who have one of their highest three elevations on DFMelan but not significant elevation ($BR < 75$) have traits that are considered dejected. These individuals are self-deprecating and apathetic and feel hopeless and discouraged much of the time. Some are philosophically tortured and dejected, while others are more moody, irritable, and dejected. Most present themselves in a way that elicits sympathy and support from

others around them. Feedback to these individuals may focus on the generally pessimistic “tint” with which they view life, as well as the hardships they have had to face.

Forlorn

Individuals who have one of their highest three elevations on DFMelan with a moderate elevation have traits that are considered forlorn. Their self-deprecation and pessimistic outlook makes them so mooney that those individuals who try to offer support and empathy are likely to give up, realizing that their efforts are in vain. This resulting lack of support simply serves to reinforce the negative feelings and pessimism, which these individuals respond to either by deepening their gloom, guilt, and self-condemnation or by presenting in a sour, irritable, and resentful manner along with self-pity. Feedback to these individuals may focus on the depth of their sadness and how it can push others away, so that they do not have the support they need.

Melancholic

The melancholic personality style involves not merely recurrent symptoms of depression but an enduring pattern of thoughts, attitudes, behaviors, and self-concepts related to depression. These clients perceive themselves as worthless, vulnerable, inadequate, unsuccessful, and guilty, and they frequently engage in self-criticism. When possible, they frame events in a defeatist, fatalistic manner. They have learned to expect ridicule and derision. Even extremely slight signs of indifference might be interpreted as contempt and condemnation. Others perceive them as forlorn, somber, discouraged, and hopeless. They similarly describe themselves as discouraged, quiet, drained of energy, and despairing. Initially, their depressive behavior might elicit support and empathy from well-intentioned others. Eventually, however, they end up feeling deserted and abandoned because their interpersonal behavior is likely either to distance others or to attract persons who will use their passivity and depression to exploit or otherwise control them. They rarely engage in active, assertive behavior to obtain reinforcement from others. They feel powerless and at the control of forces beyond their control. Although they crave love and support, they fail to act in ways that others find attractive and gratifying. Sometimes their self-criticism is a tactic to diffuse the potential criticism of others and simultaneously solicit support and sympathy. As a result, their interpersonal style serves to further reinforce their depression, and they frequently end up feeling angry, resentful, and pessimistic. Depressive/melancholic personalities can be distinguished from major affective disorder and dysthymia in that, with a melancholic personality, there will be an early, extended onset (versus more rapid and intense), along with multiple personality traits consistent with depression.

Because melancholic individuals are quite introspective, they have the potential for and the orientation toward developing depth of insight. In addition, they are emotionally responsive and often have depth of feeling. Their level of distress may also be used as an aid in motivating them to change. High-functioning melancholic individuals may be able to have genuinely close, caring relationships with others and may be articulate, conscientious, responsible, and insightful. They might potentially respond well to humor, elicit liking in others, and be able to effectively take into account alternative points of view.

Frequent Code Types

The most likely elevations on the clinical scales would be on Persistent Depression, Major Depression, and, possibly, Bipolar Spectrum. Elevations on these scales would be natural extensions of the individual's overall melancholic style. Considerable conceptual and clinical overlap with other personality scales is likely, resulting in frequent associated elevations on Schizoid, Avoidant, Negativistic, Masochistic, and Borderline. An associated elevation on Schizoid would amplify the individual's apathy and introduce an indifferent element to the melancholic style. Because these clients are more likely not to be interested in interpersonal relationships, developing an effective therapeutic working relationship would be difficult. Organizing and logically communicating thoughts is often extremely difficult. If Melancholic and Avoidant are both elevated, the melancholic style is characterized by anxiety and fear of interpersonal humiliation, which leads to isolation as they attempt to protect themselves. Such persons engage in extreme introspection (mostly of a negative nature) and have a sense of alienation from themselves. They are inhibited, have few social skills, feel easily embarrassed, and have few close friendships. They are also likely to have difficulty experiencing pleasure and feel inhibited about pursuing goals. An elevation on Negativistic flavors the melancholic personality with anger, irritability, and sour grumbling. These individuals vacillate between being bitter and resentful toward others and being intropunitive and self-deprecatory. Because they are uncomfortable with their anger and resentment, these feelings are typically expressed in indirect ways, such as through obstinacy, procrastination, and inefficiency. There are clear similarities between Melancholic and Masochistic traits. Both of these scales emphasize behaviors that result in the person's not obtaining what he or she wants from life. However, elevation in both of these scales highlights active maneuvers that result in possibly undeserved blame and unjust criticism. These persons present themselves as self-effacing, self-sacrificing, obsequious, and deserving of painful consequences. They are likely to get drawn into relationships that are physically or emotionally abusive, and they respond by being ingratiating and submissive. A Borderline and Melancholic configuration emphasizes a serious difficulty with controlling affect and behavior. Cyclical variations of emotional constraint and criticism are followed by impulsive outbursts, sometimes of a self-destructive nature. Suicide potential needs to be monitored carefully. These individuals are likely to have difficulty comforting themselves when distressed and feel that life is meaningless. Problems are typically expanded out of proportion (catastrophized). Accusations may be made that others have mistreated them. Their level of self-identity is extremely weak, and sometimes they have difficulty logically organizing their thoughts and emotions.

Treatment Implications

The major focus of intervention should be to work with their sense of helpless immobility and their belief that emotional pain is an inevitable life condition. Interventions related to interpersonal behavior, cognitive schemas, self-concept, and expectations are often essential. Specific techniques might include social skills and assertiveness training, cognitive interventions that challenge underlying assumptions, behavioral programs that enhance pleasure-related activities, and group involvement that combines

support and encouragement for change. Initial contact should be characterized by support that seeks to satisfy some of the client's dependency needs without fostering further helplessness. Psychopharmacology might be considered but should not be an end in itself. Long-standing cognitions, modes of interacting with others, and self-concept persist even after medication might have removed some of the more symptomatic features of the disorder. Therapeutic challenges involve preventing self-harm, preventing the client from proceeding too fast and possibly encountering failure and disillusionment, and preventing relapse. Relapse prevention can be enhanced by realistically advising that some recurrent difficulties are inevitable.

Scale 3: Deferential-Attached-Dependent (DADepn)

The core characteristics measured by this scale are related to levels of independence, capacity to confidently make decisions, and comfort with being assertive. Not only in relation to interpersonal relationships, this scale reflects a person's ability to care for him- or herself. Primarily, though, interpersonal dynamics are evaluated on this scale, including the degree of need for interpersonal closeness and support.

The *Puerile Expressive Emotion* facet scale reflects the degree to which individuals are confident as adults with independent perspectives and voices. Individuals who score high on this facet scale lack this confidence in taking on adult responsibilities and behave in interpersonally obsequious ways. They are overly cooperative with others, often to the point that they do not get their own needs met or their own perspective heard. Their passivity can easily become helplessness and neediness toward others, and they tend not to express dissenting thoughts or feelings. Feedback to such individuals may focus on their lack of confidence in themselves making adult decisions.

The *Submissive Interpersonal Conduct* facet scale reflects the hallmark traits of the DADepn scale related to autonomy versus subordination within interpersonal relationships. Individuals high on this facet scale abdicate their own personality in order to let others make decisions for them, including avoiding any hint of asserting their own opinions and feelings, denying any differences between themselves and others, and ultimately being submissive and complying with the will of others. They behave as helpless individuals who desperately need others to help them make even small decisions. Feedback to such individuals may focus on the intense fear that they may be abandoned to care for themselves and have to make decisions that they do not feel equipped to make.

The *Inept Self-Image* facet scale reflects the degree to which individuals see themselves as competent versus incompetent. Individuals who score high on this facet scale view themselves as weak, inadequate, and incompetent, unable to make decisions and otherwise accomplish things on their own. They feel vulnerable when they are alone or feel abandoned, minimizing or even denying any competencies they do have. They have so little faith in themselves that they often do not trust their own beliefs and opinions. Feedback to such individuals may focus on these feelings that they are not capable of accomplishing things on their own.

Deferential

Individuals who have one of their highest three elevations on the DADepn but not significant elevation ($BR < 75$) have traits that are considered deferential. These individuals outwardly exhibit low self-esteem and are described as devoted,

accommodating, and agreeable. They are also, however, good at empathizing with and caring for others and can maintain deep, lasting friendships and relationships. They are both eager to please others and easy to please themselves. Feedback to these individuals may focus on their tendency at times to be overly agreeable and accommodating to others as well as their ability to be devoted, loyal, and loving.

Attached

Individuals who have one of their highest three elevations on DADepn with a moderate elevation have traits that are considered attached. Beyond just interpersonally agreeable and accommodating, these individuals tend to be truly submissive to others, as well as inept in their own independent, adult capacities. Some become so attached to others that they lose their own individual personality and identity, gaining their purpose in life and defining who they are through their significant relationship with the other person. Feedback to these individuals may focus on their tendency to give themselves over to others and place a great deal of importance on their attachment to other people.

Dependent

The core characteristic for persons with elevations on this profile is their feeling incapable and incompetent of functioning independently and, therefore, unable to create strong bonds with people whom they perceive as being able to lead and care for them. They quickly create alliances and give up responsibility for decisions. Thus, they feel inadequate and insecure, and they have low self-esteem. They usually describe themselves as placating, insecure, passive, immature, and deserted. A primary way in which they deal with these feelings is to identify with stronger people and define themselves in terms of these people. They are continually concerned with the possibility of losing friends. To maintain friendships, they are extremely submissive and cooperative, and they cover up any unpleasant emotions out of fear that the emotions might alienate others. They, therefore, minimize objective problems, rarely disagree with others, and never take a strong position on an issue. Others, therefore, perceive them as gullible, wishy-washy, humble, timid, docile, and passive. Internally, they have a limited range of competencies in reducing tension and stressors. Elevations on this scale are consistent with bulimia (check Noteworthy Responses related to eating disorders: items 69, 86, 102, 186).

Often, dependent personalities are well liked because they are cooperative, compliant, and humble, and they value the opinions of others. They are also likely to be loyal, warm, tender, and noncompetitive. They attempt to develop and maintain lasting friendships and do so in part by defusing unnecessary conflict.

Frequent Code Types

The most frequent clinical syndrome likely to be elevated along with DADepn/Dependent is generalized anxiety, which might include panic attacks, social phobias, and agoraphobic attacks often related to or triggered by fears of separation. Mood disorders are represented by associated elevations on Persistent Depression as well

as Bipolar Spectrum and Major Depression, which are also common. Frequent associated scale elevations on the personality scales include SRAvoid/Avoidant (see the section titled “Scale 2A: Shy-Reticient-Avoidant (SRAvoid)”) and AAMasoc/Masochistic, with less frequent associated elevations on SPHistr/Histrionic, DFMelan/Melancholic (see the section titled “Scale 2B: Dejected-Forlorn-Melancholic (DFMelan)”), RCComp/Compulsive, and DRNegat/Negativistic. When Masochistic is elevated with Dependent, it highlights these clients’ poor self-esteem, based in part on having been in a series of relationships that have been painful. Although they desperately want others to care for them, they present themselves in a negative and pessimistic manner. Eventually, they undermine and sabotage the relationships that, on another level, they seek to create. High scores on Histrionic indicate that these clients are active and outgoing in attempting to get others to notice and take care of them. To this end, they might appear charming, dramatic, seductive, and extroverted, even though they would not assert themselves in a way that would highlight differences or confrontation. They are often quite sensitive to the moods of others but may have noteworthy difficulty and a feeling of emptiness when they have to act independently. An associated elevation on Compulsive indicates that dependent characteristics are combined with seeking approval and nurturance from others by acting perfectionistic, disciplined, orderly, industrious, and persistent. They are highly respectful and even ingratiating toward not only persons in positions of authority but pretty generally most other people. They will make careful preparation for future events. Their difficulty making decisions is a result of both their dependency and their focus on details. Finally, elevations on Negativistic along with Dependent indicate that although these persons seek the guidance and leadership of others, they are also quite conflicted about these relationships. They may vacillate between appearing to cooperate and then feeling resentful and angry, which leads to resistance toward others in power. Guilt follows, but then the cycle is likely to repeat itself.

Treatment Implications

Dependent personalities frequently seek treatment. Typically, rapport is quite easily established, especially if the therapist responds in an authoritative, comforting, and assertive manner. However, the greatest danger (or challenge) is that a relationship may be created in which the therapist becomes a rescuer, thereby reinforcing the dependent pattern. These clients may prefer the therapist to be directive, but a nondirective, Socratic method is more likely to encourage assertion and independence. However, a balance must be struck in order to discourage premature termination. An important goal is to reduce their clinging patterns and instead encourage their interacting in a more direct, assertive manner. Specific techniques might include assertiveness training, anxiety reduction skills (e.g., deep breathing, muscle relaxation, meditation, self-hypnosis), role playing, group therapy (to explore their impact on others), and psychoanalytic techniques that can probe the origins of their dependent patterns.

Scale 4A: Sociable-Pleasuring-Histrionic (SPHistr)

The core characteristics measured by this scale are related to the degree to which individuals think and behave in ways that convey a need for attention. Much of what is measured on this scale is interpersonally behavioral, including seductive and dramatic

behavior, wild behavioral displays to garner attention, and theatricality. However, some characteristics are cognitive and emotional, such as easily changing emotions and high suggestibility.

The *Dramatic Expressive Emotion* facet scale reflects the degree to which individuals have rapidly changing, short-lived emotional states and a tendency to emotionally overreact in dramatic and provocative ways. Individuals who score high on this facet scale display shallow but theatrical emotions, with the result of drawing attention to themselves. Their emotional displays often make them the center of attention, even when the situation is not actually focused on them. They have difficulty tolerating boredom and inactivity, which leads to impulsive and acting-out behaviors. They thrive on momentary thrills and excitements and have difficulty delaying gratification. Feedback to such individuals may focus on their emotions fluctuating at times wildly and dramatically.

The *Attention-Seeking Interpersonal Conduct* facet scale reflects the tendency to behave in ways that manipulate situations and other people into giving them praise, approval, and attention. Individuals who score high on this facet scale are socially ingratiating; that is, their behaviors initially appear to be friendly, cooperative, charming, complimentary, and flirtatious. However, they are much better at being charming and fun than they are at forging meaningful relationships. Their interpersonal behaviors are shallow and most often focused on remaining the center of attention, and they are adept at manipulating others to meet their needs and selling themselves in different situations, reading others for what is likely to be appreciated in any moment and get them positive feedback. Feedback to such individuals may include their adeptness within social situations, as well as their tendency to want to be the center of attention.

The *Fickle Mood/Temperament* facet scale reflects how rapidly shifting and shallow versus stable and deep an individual's mood states are. Individuals who score high on this facet scale are highly emotionally reactive, even emotionally erratic, experiencing emotions and displaying affects freely, dramatically, and variably. Although they are full of life and often exciting, they are just as easily angered as they are excited and enthusiastic. Their emotions can be fleeting and ephemeral, though; they do not tend to be deep and long-held. Feedback to such individuals may focus on their intense and variable moods, as well as how quickly and easily they react to emotional stimuli.

Sociable

Individuals who have one of their highest three elevations on the SPHistr but not significant elevation ($BR < 75$) have traits that are considered sociable. These individuals go out of their way to be charming, to appease others, and to be appealing socially. While they may get quite a bit of positive feedback and reinforcement for these behaviors, they often change who they are based on the social situation, revealing the possibility that they lack a strong core or inner sense of identity. Their identity is largely based on the approval and attention they receive from others. Feedback to these individuals may focus on their extremely sociable nature, as well as their strong emotional reaction to attention and approval from others.

Pleasuring

Individuals who have one of their highest three elevations on SPHistr with a moderate elevation have traits that are considered pleasuring. Also sociable in their behavior and

often charming, these individuals step up the behaviors that please others and gain attention to a degree that begins to feel false and manipulative to others and makes these pleasuring individuals frantic and cognitively scattered. Some become overly theatrical and dramatic in their efforts to gain attention and approval, including being overly sexualized, dramatically shifting their personalities to fit different situations and people, and presenting themselves in a striking manner, through flashy clothes and dramatic expressions. Others are more anxious and regressive in their behaviors aimed at eliciting attention, including demanding and clingy behavior, blatant and inappropriate sexual provocativeness, and moodiness and pouting. Feedback to these individuals may focus on the efforts they go to in order to elicit positive feedback from others and how deeply it hurts and disorients them when they feel they do not get this positive feedback.

Histrionic

Histrionic persons are dramatic, colorful, and emotional. Their tolerance for boredom is extremely low, and they are constantly seeking new situations. By focusing on the external world, they do not fully digest and integrate their experiences with their inner world. Because experiences are not integrated, they do not grow and learn from them. As a result, their level of maturation does not progress. They typically become highly invested in situations or with friends, but, when the excitement ends, they reinvest their energy and interest elsewhere. They typically describe themselves as active, egocentric, exhibitionistic, flighty, extroverted, and flirtatious. They also see themselves as charming, outgoing, and able to acquire the attention of other people. As a result, they make very good impressions in party-type situations, although sometimes they might be perceived as too loud, demanding, and uncontrollable. In addition, they might be exhibitionistic and seductive, placing excessive reliance on physical appearance. Because they react easily and spontaneously to new situations, it is easy for them to mingle with people and quickly establish friendships. However, behind these seemingly assertive and independent behaviors are strong needs for dependency. Whereas dependent personalities seek the protection and guidance of others, histrionic personalities also need the attention and support of others but seek it in an extroverted, overt manner rather than using more submissive methods. Behind histrionic personalities' dramatics and high level of activity are often conflicted, painful feelings that they avoid focusing on. Thus, their activity allows them to skim the surface of these feelings. Dissociative techniques, including the development of conversion reactions, may even be used. Typically, they communicate in a global, vague manner in which they make arbitrary judgments with little focus on the specifics of an event or concept.

Histrionic personalities can be warm, colorful, interesting, engaging, and emotionally responsive; typically, they have a good sense of humor. They adapt to new situations easily and, at least superficially, appear to have little difficulty interacting with and becoming close with others. Those who fall in the Histrionic range on this scale tend to have an above-average number of positive life events and low levels of distress, which may be adaptive.

Frequent Code Types

Because of their underlying feelings of dependency, histrionic personalities are likely to experience separation anxieties or, as an expression of their fears of emptiness, agoraphobia (check Generalized Anxiety). Conversion symptoms or hypochondriasis might also be a means of dramatically expressing their needs (check Somatic Symptom), and their need for stimulus seeking may result in substance abuse (check Alcohol Use and Drug Use). Possible associated elevations on personality scales include DADepn/Dependent (see section on DADepn/Dependent), CENarc/Narcissistic, DRNegat/Negativistic, ADAntis/Antisocial, and RCComp/Compulsive. Elevations on Somatic Symptom might indicate conversions. An elevated Narcissistic scale frequently occurs with and is quite consistent with Histrionic in that it exaggerates many of the self-centered qualities of the histrionic personality. Such clients are also likely to emphasize how charming and capable they are and to belittle those who do not partake in reinforcing their own sense of self-importance. Their descriptions of their competence and exploits are often exaggerated. They continually indicate how they are special and worthy of more attention and praise than others. An associated elevation on Negativistic is problematic in that the individuals do not like to accept their own negative emotions, such as anger and resentment. As a result of this conflict, they are moody, unpredictable, and emotionally reactive. They might overtly criticize or show disdain for others or, in contrast, express these feelings in a more indirect way, such as through obstructionism. Their attempts to repress and overcontrol their anger and resentment may sometimes culminate in explosive outbursts, followed by guilt and apologies. Similarly, an elevation on Antisocial creates conflict for these persons. They are highly dependent on others, but they also realize that their anger, disaffiliation, and resentment are likely to distance the very people whom they so much need. They might begin a relationship by being charming, friendly, and engaging, but eventually their antisocial feelings become expressed in resentment, mistrust, and even anger. In extreme cases, they might fluctuate between overcontrol and occasional extreme emotional or even physical outbursts. They may also seek to cope with this conflict through passive-aggressive strategies. They perceive their world as a competitive, potentially dangerous place, and, given these perceptions, they have become competitive, tough realists who believe that this is the only means of coping. Elevations on Compulsive also present a conflicted relationship because part of the person wants to be unrestrained and emotional whereas another part believes in the importance of emotional overcontrol. These clients are likely to seek approval through being orderly, efficient, dependable, and by dressing correctly. Often they have difficulty integrating these two modes of adapting and may become tense and moody, and their emotions may spill out inappropriately at particular moments.

Treatment Implications

Histrionic personalities are typically motivated to come to therapy because they have been through a time when they have been criticized and feel socially deprived. They describe feeling empty, bored, lonely, and discontented. Because they are emotional, responsive, friendly, and seek the support and approval of others, they are likely to

become easily engaged in therapy. These qualities usually lead to an initial high level of motivation and a good prognosis. They are unlikely to develop severe or chronic forms of psychopathology. However, they usually stay in therapy only long enough to become stabilized and rarely engage in deeper levels of self-exploration. One of the primary goals is to reduce their overdramatization. A calm, objective, cognitive approach is often useful in achieving this goal. In addition, group or family interventions can be useful in enhancing and practicing improved interpersonal skills. Given their externalizing coping style, a behavioral approach, combined with the development of specific skills, is likely to be more effective than one attempting to develop extensive insight.

Scale 4B: Ebullient-Exuberant-Turbulent (EETurbu)

The core characteristics measured by this scale are vitality in engaging with the world and the global level of joy and other positive emotion that characterize the typical experience of individuals. Seemingly overwhelmingly positive, both the vitality and globally positive emotion of these individuals can begin to clutter the mind and erode rational, logical, and realistic perception and thought.

The *Impetuous Expressive Emotion* facet scale reflects the level of boisterousness and emotional excitability an individual experiences. Individuals who score high on this facet scale are tirelessly excitable and emotionally unrestrained, which can take the form of manic energy or hot-headedness. They are socially intrusive, overenthusiastic, and “in your face.” They wake up with vigor and make themselves busy, rapid in their thoughts and actions because of their brimming with energy. This may not translate into an actual increase in productivity, though, as their energy may in fact get in the way of them accomplishing things. They may jump from activity to activity in their manic energy or may simply confuse themselves on tasks that require steady care. Feedback to such individuals may focus on their limitless amount of energy, which can be adaptive in terms of fully engaging with life but can also be detrimental as it can get in the way of focusing and persisting.

The *High-Spirited Interpersonal Conduct* facet scale reflects the energy and manner in which individuals engage socially with others. Individuals high on this facet scale have a social vitality and vivaciousness that is quite engaging and lively. They approach others with great energy and verve, and their exuberance can be contagious in a group setting. However, at higher levels on this facet scale, individuals become pushy and unrelenting, with others often experiencing them as overbearing. Feedback to such individuals may focus on the high spirit with which they engage in socializing, as well as the potential costs for that energy sometimes being too great.

The *Exalted Self-Image* facet scale reflects the degree to which individuals see themselves as special and as a dynamic force in life. Individuals who score high on this facet scale view themselves in a grand way, exalting their own energy and the vitality with which they approach life. Their self-esteem is hard to shake and is consistently high, feeling that they are popular and extremely likable. At extreme scores, this unwaveringly positive self-image can easily become grandiosity. Feedback to such individuals may focus on their perception of themselves as outgoing, charming, and full of energy. Additionally, how realistic this self-view is may be a focus of feedback.

Ebullient

Individuals who have one of their highest three elevations on the EETurbu but not significant elevation ($BR < 75$) have traits that are considered ebullient. These individuals are charming, energetic, and clever and witty thinkers. They tend to be thrill-seekers who have difficulty delaying gratification, acting impulsively and not thoroughly considering the potential consequences of their actions. They are upbeat and approach life and its opportunities with vim and vigor, thinking and behaving quite quickly. They are vivacious and can be quite creative and innovative, often quite good at solving innovative problems or solving conventional problems in innovative ways. Feedback to these individuals may focus on the high energy with which they attack life and its opportunities, including the potential problems associated with impulsive and impetuous behavior.

Exuberant

Individuals who have one of their highest three elevations on EETurbu with a moderate elevation have traits that are considered exuberant. These individuals have all the energy, vitality, and vigor with which ebullient individuals approach life, but the energy of exuberant personalities becomes too high and their behaviors too impetuous. They are interpersonally intrusive, as their energy and impatience for frustration tolerance make them irritable when they do not feel they are getting the right amount of energy and positive feedback from others. They are too preoccupied by their own energetic vitality to consider the feelings of others during interactions. Additionally, their rapid minds, while still clever and creative, become disorganized and scattered. Some go to great lengths to seek out constantly pleasurable activities, abandoning any situation that does not give them instant gratification. Others channel their high energy toward grand schemes and entrepreneurial endeavors, though rarely successfully, as their manic energy leads them to abandon ideas for other “great opportunities” that come along. Feedback to these individuals may focus on how their incredible energy level can help them succeed in many areas of life, but it can also get in the way of a great number of different areas of being productive and interacting with others.

Turbulent

Turbulent persons are reckless, erratic, and manically energetic. Their tolerance for boredom is extremely low, and they are constantly seeking new and pleasurable situations. Their baseline energy level is so high that their moods are intense and can shift quite quickly. They are typically reckless in their behaviors, impulsively and impetuously seeking out pleasurable experiences and manically engaging in careless, ill-advised schemes, which rarely (if ever) work out. They typically describe themselves as active, energetic, vivacious, full of life, and high-spirited. They also see themselves as charming, outgoing, and exciting. Like histrionic personalities, they make very good impressions in party-type situations, although sometimes they are perceived as pushy, intrusive, and overbearing. In addition, they might be on the verge of or on the other side of a nervous breakdown, having pushed themselves beyond what they can handle. Because they tend to engage the world with 100% of their energy and

resources, their wild and reckless behavior may have caught up to them, depleting them emotionally and physically. What results is a collapse into exhausted despair and depression. Unlike the erratic emotions of histrionic personalities that are shallow, while the emotions of turbulent personalities may fluctuate rapidly, they are deeply and seriously felt, from euphoria to despair.

Turbulent personalities can be engaging, enthusiastically supportive, and extremely fun and fun-loving. Their energy can be contagious, and while their overbearingness can get in the way, they generally have little difficulty interacting with and becoming close with others. The vigor with which they attack life is exciting and pleasurable, dampened only when they exhaust themselves.

Frequent Code Types

The most frequent clinical syndromes likely to be elevated along with EETurbu/Turbulent are mood disorders, which might include Bipolar Disorder and depression. The Persistent Depression and especially Major Depression scales should be examined to determine whether and how well the individual's exuberant mood is containing underlying depressive feelings. Frequent associated scale elevations on the personality scales include CENarc/Narcissistic and SPHistr/Histrionic, with less frequent associated elevations on ADAntis/Antisocial, DRNegat/Negativistic, and MPParaph/Paranoid. When Narcissistic is elevated with Turbulent, it highlights the grandiose nature of their self-appraisal. These individuals need constant praise for their accomplishments and have boundless energy to intrusively push for this praise. They often exaggerate or even fabricate accomplishments, which can appear outlandish, in order to appear praiseworthy to others. High scores on Histrionic indicate that these clients are extremely active and energetic in attempting to get others to notice and take care of them. To this end, while they might appear charming, dramatic, seductive, and extroverted, they have exaggerated social intrusiveness and overstep social boundaries frequently. Their mood states are particularly erratic. An associated elevation on Antisocial indicates that their recklessness gets to the point of violating norms and rules. Their pleasure- and stimulation-seeking is especially highlighted, as they chase highs regardless of consequences. They may be particularly prone to abusing alcohol and especially drugs (check Alcohol Use and Drug Use). Elevations on Negativistic along with Turbulent indicate that feelings of resentment can escalate easily into outbursts, though these may not be directly aggressive toward the intended target. The lack of ability to tolerate frustration and delay will drive the negativistic personality to abandon situations that cause resentment. Finally, elevations on Paranoid along with Turbulent reflect an active, hypervigilant, truly paranoid, possibly manic state. Elevations on Bipolar Spectrum and especially Delusional should be considered carefully with this combination, as the paranoid thoughts can drive the turbulent personality to drastic behaviors.

Treatment Implications

Turbulent personalities are typically motivated to come to therapy either at a moment when they are exhausted and swinging more toward a melancholic state or because their impulsive energy has gotten them into some kind of trouble. Those characterized

by some sadness and fatigue are more easily engaged in treatment, as they are uncomfortable in their current state. Those who have sought out (or been mandated to) treatment who are not in an uncomfortable state have multiple resistances to effective treatment. Their inflated self-esteem can block them from truly engaging in treatment, as they feel that they do not truly need to change. As such, care should be taken not to challenge or outwardly judge them too harshly. Additionally, their heightened and fluctuating mood states require therapists to be on their toes, as these individuals may react differently to different stimuli at any given moment. One of the primary goals in treatment is to increase their capacity to self-control. An empathic approach is key, as the therapeutic relationship is important and not necessarily easily forged. As with histrionic personalities, given the externalizing coping style, a behavioral approach, combined with the development of specific social and self-management skills, is likely to be more effective than one attempting to develop extensive insight. Finally, medication treatment should be considered if the turbulent personality is significantly impairing functioning.

Scale 5: Confident-Egotistic-Narcissistic (CENarc)

The core characteristics measured by this scale are related to the degree to which individuals see themselves not only as worthy of praise and admiration, but also simply as the center of their world. In general, the degree to which people see themselves as central to their own existence and lives varies inversely with how much regard they have for others in their lives.

The *Exploitive Interpersonal Conduct* facet scale reflects the level of self-centeredness and conversely reciprocal relationships and interactions in which the individual engages. Individuals who score high on this facet scale are entitled and expect others to praise them, admire them, and give them special favors. They do not, however, engage in these behaviors reciprocally; they can be callous, indifferent, and dismissive of others, as they feel they themselves are more special than others around them. This unempathic stance easily becomes manipulative and exploitive, using others to meet their needs and enhance themselves. It is important to note that there are some of these individuals who tend to indulge these entitled interpersonal behaviors (notably those who also score high on DADepn/Dependent), though most individuals do not respond well to them. Feedback to such individuals may focus on the fact that they feel personally entitled to good fortune and admiration by others and that they feel other people should be grateful to have relationships with them. Further feedback may focus on their lack of empathy toward others and the potential consequences of it.

The *Expansive Cognitive Style* facet scale reflects the degree to which individuals' thoughts and thinking processes are constrained or not by reality. Individuals who score high on this facet scale have thought processes that are not highly constrained by reality. They tend to exaggerate accomplishments, minimize failures, or even reconstitute failures into successes in some way. They may not simply embellish but in fact lie in order to reinforce their view of themselves as special, amazing, and worthy of admiration. Their cognitive style often centers on justification. They justify their own high self-worth with any evidence (real, exaggerated, or made up), and they justify

(rationalize) any failures as the fault of others. They are quick to denigrate others who do not quickly and easily reinforce their high self-worth. Feedback to such individuals may focus on the style of thinking that is constantly looking for evidence of their own greatness, but often ignores any information that may lead them to question their view of themselves.

The *Admirable Self-Image* facet scale reflects the degree to which individuals see themselves as special, unique, superior to others, and ultimately worthy of admiration and even envy. Individuals who score high on this facet scale are often seen as arrogant and egotistical by others, as they feel they occupy some special place in the world that others around them do not. Most often, this expectation to be seen as meritorious is not backed up by actual accomplishment. They simply have a relatively fixed, stable view of themselves that reflects high self-worth. Feedback to such individuals may focus on their strong belief that they are extraordinary and special and that many other people do not understand just how special they think they are.

Confident

Individuals who have one of their highest three elevations on the CENarc but not significant elevation ($BR < 75$) have traits that are considered confident. These individuals are interpersonally bold because of their own self-assuredness. They are confident and assertive in their opinions, typically not needing time to be thoughtful or reflect, and they expect that other people will acknowledge and go along with them. They can be quite effective leaders and are genuinely driven to succeed, though they are also entitled and feel a sense that others should simply understand that they are great and follow them. They are resourceful and optimistic, enjoying themselves easily. They can also be quite driven in endeavors, keeping busy and conscientiously attacking projects and tending to accomplish them on their own, without the input, feedback, or collaboration of others. Feedback to these individuals may focus on their ambition and drive to succeed in ways that will show others how good they are, which aligns well with their own high self-confidence.

Egotistic

Individuals who have one of their highest three elevations on CENarc with a moderate elevation have traits that are considered egotistic. These individuals' displayed confidence is exaggerated to the point of arrogance and entitlement, which actually betrays underlying feelings of inadequacy and low self-worth. That is, these behaviors are exaggerated and put on display in order to actively elicit admiration, adoration, and envy from others, a need they have only because they do not have healthy intrinsic self-worth. They are often more invested in how they are seen by others than in being genuine or presenting themselves as they actually are. They deliberately (although perhaps not necessarily consciously) select interpersonal behaviors that are aligned with how they want to be perceived by others. They can also exploit others for their own gratification, such as engaging others in romantic or sexual teasing and temptation, in order to elicit the admiration of others (thus feeling more attractive and wanted), but without necessarily following through with the romantic or sexual interactions. Their needs for bolstering their own self-esteem render other people as purely instrumental,

ways of feeding their egos, and as such these individuals are callous and unempathic, indifferent to the effect they are having on other people. Feedback to these individuals may focus on the behaviors that serve to enhance their own self-image but may alienate or hurt other people.

Narcissistic

The central characteristic of individuals with elevations on this scale is their exaggerated sense of self-importance and competence. Because they perceive themselves as special, they are likely to assume that many of the conventional rules of living with people do not apply to themselves. In addition, they may feel that they deserve special favors without having to reciprocate the time and resources that are given to them. Internally, they might be quite creative in developing plausible reasons for their self-centeredness, but, to others, these reasons might seem flimsy and transparent. Their fantasies typically involve immature, self-glorifying situations in which they are the center of attention because they are beloved, admired, successful, and physically attractive. In real life, failures are quickly rationalized and conflicts are minimized, and they are adept at enhancing their sense of pride. In building their image, they might depreciate the value of others to make themselves look superior by comparison. They might, therefore, appear arrogant, haughty, snobbish, pretentious, and conceited. They present themselves as intelligent, sophisticated, outgoing, and charming, with an air of cool optimism and feigned tranquility. Rarely do they express any self-doubt. Interpersonally, they are likely to be exploitive, autocratic, and insensitive to the needs and feelings of others. Thus, they are generally lacking in empathy. They constantly attempt to obtain admiration from others. If they are in situations in which they are criticized, they might become quite competitive and aggressive toward those who criticize them, or they may react with contempt or indifference. Thus, they have a primarily externalizing coping style. If their narcissistic bubble is burst, they are at risk for becoming depressed and potentially involved in substance abuse. A subgroup of high scorers is well adjusted and does not experience much emotional distress. As such, for these individuals, high scores should be interpreted as merely a style of adapting rather than a possible disorder. In contrast, others are pathologically narcissistic. Thus, a diagnostic challenge is determining in which of the two groups the client best fits. Generally higher elevations suggest a more problematic style.

These individuals frequently make excellent first impressions and might even receive respect and affection from others. Typically, they are articulate, carry themselves with dignity, and have a good sense of humor. Others often perceive them as being proud, independent, confident, and optimistic.

Frequent Code Types

Because narcissistic persons are prone to affective disorders and substance abuse, check relevant clinical scales (Bipolar Spectrum, Persistent Depression, Alcohol Use, Drug Use, Major Depression). Additionally, as externalizing copers who are generally not emotionally self-aware, they are likely to exhibit emotional problems more somatically (check Somatic Symptom). Personality scales that are most likely to be elevated include SPHistr/Histrionic (see section on SHPistr/Histrionic) and

ADAntis/Antisocial. Additional elevations are often found on DRNegat/Negativistic and ADSadis/Sadistic. High scores on both Narcissistic and Antisocial suggest that these individuals feel so particularly special that the rules and laws of society should not and do not apply to them. Their exploitation of others becomes so extreme that it actually intrudes on their rights, and they show very little, if any, regard for the safety and well-being of others, justifying their rule-, law-, and norm-breaking behaviors as acceptable given their special station in life. The combination of Negativistic and Narcissistic places these persons in a difficult, conflicted position. They seek to perceive themselves as superior and special in relation to others, but they are also acutely aware of their limitations. Thus, they are likely to be apologetic, submissive, compliant, and cooperative on one hand but also hypersensitive, moody, resentful, and angry on the other. They have marked difficulty in accepting criticism, combined with frequent mood changes. Elevations on Narcissistic and Sadistic emphasize the self-centered, competitive, and possibly aggressive and intimidating character of these persons. They are likely to be hostile and exploitive and justify this conduct by pointing out the competitive and exploitive nature of other people. At times, they might become malicious, cruel, and abusive; at others times, they may be cheerful, gracious, and friendly. Because they fear the criticism and possible exploitiveness of others, they might frequently be guarded, resentful, and reserved.

Treatment Implications

Because attending therapy is an implicit admission of imperfections, it is unusual for narcissistic persons to initiate therapy themselves. When they do, it is usually because their narcissistic sense of superiority has been compromised through events such as divorce or loss of employment. Interpersonally, they are likely to remain aloof and often be competitive with the therapist. They might question how someone who is less talented than they are could possibly be of assistance. Alternatively, they might elevate and inflate the status of the therapist because their association with someone who is so accomplished can be used to bolster their own sense of self-esteem. The easiest tactic for returning them to their previous level of functioning is to encourage and support them in recounting their previous successes and achievements. However, this may do them a disservice in the end because they will not learn new strategies of coping and relating. A particularly useful technique might be cognitive restructuring, in which they are helped to challenge the need to be perfect and desensitized to criticism. Group and family therapy might support them in achieving more realistic and adaptive interpersonal skills. Given that they are likely to deny imperfections and resist change, either paradoxical interventions or approaches that use nondirective or self-directed techniques are likely to produce the best outcomes.

Scale 6A: Aggrandizing-Devious-Antisocial (ADAntis)

The core characteristics measured by this scale are related to how actively and singularly individuals seek out excitement and pleasurable situations, regardless of the consequences. This singular aim toward excitement, pleasure, and self-centered gain leads to the disregard for the personal rights, safety, and authority of others and limits

their ability to empathize, understand potential consequences of their behaviors, and ultimately to learn and modify their behavior.

The *Irresponsible Interpersonal Conduct* facet scale reflects how undependable, untrustworthy, and ultimately likely to violate the rights of others an individual is. Individuals who score high on this facet scale are unreliable, as their only responsibility is to themselves and their own pleasure, gain, and excitement. They are intrusive and violate the rights of others as well as disregarding their adult responsibilities. They are disrespectful and disobedient, and they often make specific gestures to upend social conventions, norms, rules, and laws. They are very likely to be deceitful when it suits their purposes and to break the law. They are also adept at lying, both in order to get out of trouble and to make themselves look good for potential future gain. Feedback to such individuals may focus on the problematic interpersonal behaviors they exhibit, as well as their lack of remorse and guilt associated with them.

The *Autonomous Self-Image* facet scale reflects the degree to which individuals feel most secure, free, and satisfied when unencumbered by personal responsibilities, loyalties, and attachments. Individuals who score high on this facet scale revel in their freedom from responsibility and burdensome attachments. They feel unconfined by other people, as well as by the rules of society, jobs, and obligations. They pride themselves on being untethered to social convention. They value their own abilities to deceive others cunningly and creatively, as well as their overall disregard for the rules of society. Feedback to such individuals may focus on the high value they place on the fact that they do not feel the need to be responsible or “answer to” others, including the fact that this freedom can hurt others and get them into trouble at times.

The *Acting-Out Intrapsychic Dynamics* facet scale reflects how constrained versus untethered individuals’ negative inner impulses are. Individuals who score high on this facet scale do not tend to constrain, hold back, repress, alter, or even delay their aggressive, manipulative, and malicious impulses. They let their impulses drive their behavior in an unrestrained manner, and they generally do not feel guilt or remorse for doing so. Part of the underlying dynamic that contributes to this unconstrained discharge of negative impulses is that these individuals tend to use projection quite often, interpreting ambiguous and subtle cues as attacks, indifference, or generally behaviors with malicious intent. This easily justifies behavioral acting out without remorse. Feedback to such individuals may focus on their tendency to act on their feelings, especially negative feelings, quite quickly and easily, without much forethought and without remorse.

Aggrandizing

Individuals who have one of their highest three elevations on the ADAntis but not significant elevation ($BR < 75$) have traits that are considered aggrandizing. These individuals are fiercely independent and choose to approach life and situations in it, including relationships, in a nonconforming way. Their major concern is for their own self-enhancement, though they approach their own enhancement in generally socially acceptable ways. They can be confrontational, assertive almost to the point of aggressive, and pushy, and they generally make fairly effective leaders. They are undeterred by obstacles that block their path and quite resourceful in meeting their needs and desires. Some are greedy, power hungry, and exploitive of others to enhance themselves and

their status. Others are actively noncompliant, oppositional, and impulsively intrusive, simply wanting to ensure that their situation, good or bad, is being driven by their own behaviors, rather than as a consequence of others' influence. Feedback to these individuals may focus on their general drive toward personal gain and the lengths they will go to in order to achieve it.

Devious

Individuals who have one of their highest three elevations on ADAntis with a moderate elevation have traits that are considered devious. These individuals focus on self-serving behaviors, but these behaviors are increasingly impulsive and irresponsible. The major means of building themselves up is through devious means, manipulating, lying, and generally making trouble for other people. These individuals are generally risk takers, impulsive and foolish (rather than courageous) and either unaware or unfazed by potential consequences. Beyond their primary focus on increasing their own perceived value, these individuals seek out excitement for excitement's sake, engaging in reckless and risky behaviors, from which they gain a feeling and sense of being free and unburdened by the "laws" of the world. In addition to material and status self-enhancement, these individuals place a great deal of value on defending their reputations for being strong, tough, and fearless. Feedback to these individuals may focus on the lengths they will go to in order to feel free and strong, as well as to ensure that others know that they are people not to be messed with.

Antisocial

The central theme for persons with elevations on this scale is competitiveness, along with impulsive acting out of antisocial feelings. They are often described as provocative, violent, vicious, self-centered, dominant, dishonest, brutal, and devious. Their actions are often hasty, shortsighted, and imprudent, and they generally ignore the consequences of their actions even to the extent of disregarding the safety of themselves and others. They can be interpersonally irresponsible—they will violate the personal rights of others in occupational, marital, parental, or financial contexts. They can be expected to have legal difficulties because many of these individuals engage in criminal activities. For others in this category, legal problems are often absent because they confine their acting out to legal domains, such as alcohol abuse, interpersonal insensitivity, unreliable work practices, and irresponsible sexual behavior. However, they do not conform to social norms and may even feel and express contempt toward these norms. They enjoy the feeling of not being confined by standard modes of conduct and project the image of being free, flexible, unencumbered, and having little obligation to schedules, commitments, or persons. This image is usually associated with a lack of compassion, empathy, remorse, and charitableness. They justify frequent expressions of callous competitiveness by pointing out the exploitiveness of others or otherwise conceptualizing the world as functioning according to the law of the jungle. Because of these attitudes, they are mistrustful, suspicious, guarded, and reserved. They might also be aggressive, intimidating, cold, insensitive, or even cruel and malicious, thereby provoking fear. They may treat with contempt those whom they consider "weak," or they might ascribe their own malicious tendencies to others. When challenged, they are likely to become impulsively angry or resentful, vindictive, and vengeful.

At their best, antisocial personalities can be gracious, charming, friendly, and cheerful. Some people might perceive them as interesting and exciting, at least in part, because they are not confined by the same rules of conduct and restraints as other people.

Frequent Code Types

Check to see whether the clinical scales of Alcohol Use and Drug Use are elevated; given the impulsiveness and hedonism of antisocial personalities, they are prone to substance abuse. Although generally free from anxiety, they can develop mood disorders, especially when being held accountable for antisocial acting out or feeling confined or out of control (check Bipolar Spectrum, Persistent Depression, and Major Depression). Associated personality scales that are most frequently elevated include CENarc/Narcissistic (see section on CENarc/Narcissistic) and ADSadis/Sadistic. Additional personality scales that may be elevated include SPHistr/Histrionic (see section on SPHistr/Histrionic), DRNegat/Negativistic, UBCycloph/Borderline, and MPParaph/Paranoid. High scores on Antisocial and Sadistic are noteworthy as they indicate that acting-out behaviors will be cruel, malicious, and callous. The elevation on Sadistic indicates that the expression of antisocial feelings is direct, overt, and abusive, and hurting others is reward in itself. Such persons should be treated with considerable caution. When Negativistic is high along with Antisocial, the angry, resentful characteristics of the antisocial individual are brought out; yet the same individuals may desire the closeness and warmth that could be available in relationships. However, they perceive the world as a struggle in which most situations are framed in win-lose terms. Thus, they frequently override their need for affection by becoming tough-minded, competitive, and interpersonally superficial. They might excel in individualistic activities—some competitive sports or sales positions, for example—but they would have difficulty working in situations that require loyalty and team coordination.

When Antisocial is elevated along with Borderline, it suggests that the general discomfort felt by antisocial personalities is likely due to the general turmoil of emotions of the borderline personality. These individuals will act out erratically and often in ways that do not seem to make sense, even to the individuals themselves. They will act out impulsively on their beliefs and feelings, even when these change rapidly and contradict each other from moment to moment. These individuals are at high risk for self-harm, including placing themselves in risky situations in which harming themselves seems inevitable. When Antisocial and Paranoid are both elevated, the projection of malicious intent that is present in antisocial personalities is amplified, and these individuals actively behave in reaction to the belief that others are not only out for themselves but specifically against the antisocial, paranoid individuals. Their impulsive behaviors may seem more frantic and, unencumbered by societal rules, their interpersonal acting out may seem drastic. Again, these individuals may pose specific threats to themselves or others.

Treatment Considerations

Antisocial personalities typically do not recognize the need for treatment and are most frequently referred either by the courts or because of threats from spouses that they will leave them. Once in therapy, they are likely either to openly defy therapist interventions

or to develop a facade of cooperation in the hope that they might be able to somehow exploit the situation. Therapists need to be cautious; they can potentially be conned by these clients, who would then perceive them as weak and not worthy of respect. The therapists may then run the risk of becoming angry, cynical, and punitive—and ineffective. Given that the antisocial individual's style is one of externally acting out, the most appropriate interventions are ones that are directed toward changing specific forms of behavior with clear limits: behavior modification, behavioral contracting, and external monitoring of behavior. Antisocial personalities are unlikely to be responsive to internalizing, insight-oriented interventions. In addition, because their arousal level is typically low, techniques that increase arousal, distress, or even anxiety serve to increase their level of motivation. A group context might work particularly well, because antisocial personalities are more responsive to peer influence than to authority-directed influence. However, most interventions have not been demonstrated to be effective in changing their underlying personality structure. A more realistic goal is the reduction of specific targeted symptoms or behaviors, particularly their aggression, destruction, impulsiveness, and poor affect. Target behaviors might be framed in the context that change is in the client's self-interest.

Scale 6B: Assertive-Denigrating-Sadistic (ADSadis)

The core characteristics measured by this scale are related to behaviors specifically directed toward the harm, humiliation, and pain of others, as well as the general pleasure found in inflicting these interpersonally harmful acts. Beyond the “side effect” of others suffering that is found in the ADAntis/Antisocial scale, secondary to self-enhancement, this scale focuses on the direct personal enjoyment at the deliberate infliction of harm to others.

The *Precipitate Expressive Emotions* facet scale reflects the degree to which individuals are generally unmoved by emotion and emotionally laden situations, except for sudden, reactive anger and hostility. Individuals who score high on this facet scale are generally insensitive to the feelings of others, seeing kindness, gentleness, and sympathy as weak, distasteful, or falsely masking more malicious emotions. They are cold and often appear unfeeling, but they also have difficulty tolerating frustration. When frustrated or confronted in any way, they become quickly and easily irate, reflexively aiming to demean, humiliate, or harm others in some way. They are attracted specifically to risk and potentially harmful situations, and they are unfazed by danger, pain, and punishment. Feedback to such individuals may focus on their quick temper and general suspicion of the motivation of others who display kind, sympathetic, or warm emotions to others.

The *Abrasive Interpersonal Conduct* facet scale reflects the most salient aspect of the ADSadis/Sadistic scale, the degree to which interpersonal interaction is characterized by abusive, demeaning, and intimidating behavior. Individuals who score high on this facet scale are vicious and take pleasure in intimidating and harming others. Most often they are at least verbally abusive, though some are physically and sexually aggressive as well. They work hard to dominate others across various contexts, and they freely use intimidation (including threats, sarcasm, and harsh judgment) to coerce others to submit to them. Interacting with these individuals can be intimidating and

brutal. Feedback to such individuals may focus on their frequent use of intimidation and domineering behaviors to belittle, humiliate, and coerce others to their will.

The *Eruptive Intrapsychic Architecture* facet scale reflects the degree to which generally negative, aggressive, and sexual energies that underlie a negative interpersonal outlook tend to explode into aggressive behavior. Individuals who score high on this facet scale generally feel that others are as power-hungry as they are, and the most upsetting thing would be to allow others to dominate, deceive, or humiliate them. They cannot tolerate feeling vulnerable, and these negative internal views of others breed powerful, aggressive energy. This energy is generally decently well modulated, but periodically these individuals erupt and cause the harsh, mean, and in their thinking defensively aggressive behaviors that are commonly exhibited by those high on the ADSadis/Sadistic scale. Feedback to such individuals may focus on the sudden urges (instincts or reflexes) that they feel to be harsh and cruel to others.

Assertive

Individuals who have one of their highest three elevations on the ADSadis but not significant elevation ($BR < 75$) have traits that are considered assertive. These individuals demonstrate many qualities that are sought after in leaders, including a forceful interactive style and the ability to take on a great deal of responsibility without hesitation or fear of failure. They are competitive and driven to succeed, doing whatever is necessary to ensure that others on their team behave in a way that will help them succeed. They are tough and commanding, unmoved by their own and others' emotions. Some organize their lives around humiliating and demeaning others, dominating situations whenever possible in order to satisfy a seeming quite general grudge toward others in life. Others bolster themselves by intimidating others, interrupting them, highlighting their weaknesses (even if simply by being overly assertive themselves), and ensuring that their own opinions are forcefully placed at the forefront of conversation. It is important to note that these individuals often find socially sanctioned ways to thrive as leaders, such as in politics or within high-pressure corporate environments. Feedback to these individuals may focus on their natural tendencies toward leadership, which can sometimes spill over into domineering or forceful behavior.

Denigrating

Individuals who have one of their highest three elevations on ADSadis with a moderate elevation have traits that are considered denigrating. These individuals expect to be attacked by others and, in turn, attack others first. They are rigid and hostile, interpreting any feedback as criticism. Some organize their interaction with the world around the feeling that they have a right to determine who deserves to be punished and to inflict that punishment on them. This may include actual violence and destruction. Some find socially sanctioned ways of judging and punishing others (e.g., as police officers, in the military, or as judges), whereas others simply feel empowered to "enforce" their own sense of justice outside of legally and socially sanctioned methods. A subgroup of these denigrating personalities acts out against others out of insecurity and feeling small. Finding ways as an adult to lash out at, and specifically hurt, the types of people who bullied, harmed, or otherwise denigrated them in the past, these individuals capitalize on opportunities to feel powerful. Feedback to these denigrating

individuals may focus on the aggressive and hostile behaviors they enact in order to feel powerful and in control.

Sadistic

Individuals scoring high on Sadistic are typically competitive, energetic, hard-headed, authoritarian, and socially intolerant. They are predisposed toward aggressive outbursts, which might be expressed in a callous manner with little awareness of the impact of their verbally or physically aggressive actions. In many ways, their callousness can be seen as a further pathological variation of the antisocial personality. Being in control and exerting power most often to the point of intimidating others is a central means they use to achieve their goals. Humiliating their victims also serves to release their own psychological pain. Sometimes they enter socially approved enforcing roles in which their expression of aggression is disguised behind socially sanctioned rules (the strict disciplinarian school principal or overzealous police officer). They are relatively unaffected by pain and punishment and may act in a manner that is both reckless and daring. They have a tough-minded orientation, which might be expressed in a caustic and contemptuous attitude toward social events and is consistent with their prejudice, intolerance, and authoritarianism. At their worst, they express vicious, explosive, violent, and even brutal behavior. Noticeably absent is a sense of shame, guilt, sentimentality, or internal conflict. They perceive other persons as objects to manipulate and control. This attitude might be enhanced and justified if the victims can be considered members of disempowered, marginalized groups.

A positive aspect of persons with this profile is that they can cope effectively with challenges. They can be unflinching and daring, which, if expressed in the right context, can be considered courageous. In reaching a goal, they are relatively unencumbered by subtle ambiguities that might make it difficult for other people to take action.

Frequent Code Types

Fortunately, elevations on ADSadis/Sadistic are infrequent. When they do occur, noteworthy elevations on other scales include ADAntis/Antisocial (see section on ADAntis/Antisocial), DRNegat/Negativistic, CENarc/Narcissistic, and MPParaph/Paranoid. A corresponding elevation on Negativistic suggests that the moods and behaviors of the sadistic personality may be more shifting and in flux than in those without negativistic features. These persons may fluctuate between outright aggressiveness and manipulative passive-aggressiveness in order to wield their power. When Sadistic is added to Narcissistic, these individuals not only have an inflated, unrealistic sense of themselves, but they are also likely to be openly hostile and destructive, which is not the case when Narcissistic is elevated by itself. A corresponding elevation on Paranoid indicates that these persons' cruelty might be self-justified by even stronger suspicions that others would like to exploit or even brutalize them.

In addition to these personality pattern scales, check the mood disorder scales (Bipolar Spectrum, Persistent Depression, and Major Depression). Substance abuse is also quite common among those with sadistic personalities, so also check Alcohol Use and Drug Use. Finally, hostile excitement may trigger delusional episodes, as reflected on the Delusional scale.

Treatment Implications

This difficult-to-treat group almost never reports to therapy on their own initiative. Once in therapy, they are likely to belittle the therapist and may even be overtly hostile. A therapist who responds negatively is likely to be perceived as weak, and clients use this perception to discount therapist interventions. In addition, they typically lack insight into their behavior and can even be indifferent to the damage they inflict. Cognitive interventions are unlikely to be successful because their thought patterns are quite rigid. Potentially useful approaches might be anger and impulse management programs, developing assertive as opposed to hostile communications, and persuading them to see that changing some of their more problematic behavior is actually in their own self-interest.

Scale 7: Reliable-Constricted-Compulsive (RCComp)

The core characteristics measured by this scale are related to rigidity, perfectionism, and anxiousness, all in service of meeting the desires of others in a conscientious, reliable way. Although the manifested behaviors are the focus of this scale, it also reflects the suppression of all personal impulses and desires, in order to remain oriented to tasks and accomplishment.

The *Disciplined Expressive Emotion* facet scale reflects the degree of control individuals exert over their emotional expressions. Individuals who score high on this facet scale keep tight control over their emotions, often appearing cheerless and quite humorless. Their regimented and highly organized lifestyle supports their lack of emotional freedom, and they are careful and considered in their interactions with others. Their perfectionism can lead them to be quite myopic, interfering with accomplishing tasks and making decisions, as well as interfering with the inherently messy and unpredictable nature of interpersonal relationships. Feedback to such individuals may focus on the strict control they keep over their emotions in order to remain organized, deliberate, and focused.

The *Constricted Cognitive Style* facet scale reflects the degree to which individuals organize their world according to rules, hierarchies, and rigid standards. Individuals who score high on this facet scale adhere very strongly to rules and regulations, whether formal or informal (such as social norms). They are rigid and judgmental, hard on themselves and on others whom they feel do not meet up to particular standards. They harshly judge those who are impulsive, irresponsible, and who behave “emotionally.” While they are efficient and conscientious, they lack creativity and vitality and are unable to tolerate spontaneity and unfamiliarity. They are stubborn, and their rigid thinking often gives them difficulty when making decisions. Feedback to such individuals may focus on their rigid worldview and the harsh judgment they inflict on both themselves and others for not living up to certain standards.

The *Reliable Self-Image* facet scale reflects the degree to which an individual’s self-image is bolstered by feeling that he or she is dependable, meticulous, industrious, and effective. Individuals who score high on this facet scale value the fact that they are extremely responsible, excellent workers. They most often forego leisure activities and frivolity in favor of putting extra effort into their work. They are most satisfied with themselves when they accomplish tasks to an extremely high standard

(often perfection). They fear and are extremely uncomfortable with mistakes, missed deadlines, incomplete projects, and errors in judgment in general, and as a result they value and exhibit high levels of discipline and obedience. They can be extremely moralistic and judgmental of perceived transgressions of generally accepted moral codes. Feedback to such individuals may focus on how much emphasis they place on perfection and being responsible, with the resulting unease they feel when they do not live up to their own high standards.

Reliable

Individuals who have one of their highest three elevations on the RCComp but not significant elevation ($BR < 75$) have traits that are considered reliable. These individuals are conscientious and reliable, well disciplined in their lives (work and otherwise). Though not emotionally constricted, they are more careful with their emotions and do not often act impulsively. They are organized and function very successfully in the world, driven to achieve to a high level and perform tasks completely, competently, and free of error. Mistakes motivate them to improve their performance in the future. They rarely, if ever, break rules, and they can be counted on to behave in conventional and socially acceptable ways, rarely taking significant risks. Some reliable individuals do exhibit anxiety around making mistakes and, as a result, work and rework tasks, never quite being satisfied with the result. These individuals may find their anxiety getting in the way when they are asked to accomplish a task that is novel or that they do not feel they have a firm grasp on. Feedback to these reliable individuals may focus on their dependable, conscientious nature and the possible anxiety that it may cause them at times if they do not feel they are delivering appropriately.

Constricted

Individuals who have one of their highest three elevations on RCComp with a moderate elevation have traits that are considered constricted. These individuals allow their fear of mistakes to drive them to significantly constrict their emotion and behavior. They become behaviorally and attitudinally rigid, in order to control as much of their environment as possible. This environmental control contributes to their emotional suppression, as a predictable environment rarely elicits strong emotions. Even when emotions are warranted, however, these individuals maintain strict control over their emotional reactions and expressions, as they are seen as antithetical to the reliable, conscientious task orientation that is strived for. These individuals often suffer physical (somatic) ailments as a result of the strong emotional constriction. Some of these individuals find a perfect fit in organizations that are highly bureaucratic and rule-bound. Others simply raise a self-protective wall between themselves and the outside world, especially other people, sharing little with others and retreating into conscientiousness to protect themselves. Feedback to these individuals may focus on how tightly wound they are, emotionally and behaviorally, and how they feel more comfortable and secure when they are in a highly controlled, predictable environment.

Compulsive

The core characteristics for persons with this elevation are conformity, discipline, self-restraint, and formality. They strictly adhere to social norms and may even be upset by novel ideas, especially if they challenge established norms of conduct. They are conscientious, well prepared, righteous, and meticulous, and they perform well when required to work on a schedule. They typically work hard, to the exclusion of leisure activities. Their emotions and behavior are tightly controlled. Interpersonally, they are formal, moral, perfectionistic, and rigid. They are overrespectful and even ingratiating toward persons in authority. In contrast, they are likely to be demanding, perfectionistic, and even contemptuous of subordinates, insisting that they act in strict adherence to correct and preestablished rules and methods. Self-descriptions include responsible, dependable, orderly, punctual, reliable, and stubborn. Internally, they are rigidly controlled and do not allow themselves to experience any forbidden thoughts or impulses. Their world is constructed in terms of schedules, deadlines, rules, ethics, and prescribed forms of behavior. Although they perform well in structured, concrete working environments, they have difficulty adjusting to changing work situations that require creative, spontaneous responses. These strategies provide them with a high degree of control over their world and their inner impulses, but the price they pay is a grim, tense, joyless life in which warm feelings and spontaneity are kept under tight control.

Positive qualities include loyalty, prudence, consistency, predictability, and a strong sense of duty. They are often able to approach a difficult situation with maturity and competence. In a work context, they are punctual, thorough, diligent, and honest and rarely make mistakes. Often compulsive persons are high achievers and rarely report psychiatric distress.

A defensive, fake good profile can produce an elevation on Compulsive. In these cases, the previous scale interpretation should not focus on discipline and restraint but rather on the client's defensiveness.

Frequent Code Types

Typically, elevations on RComp/Compulsive are not accompanied by elevations on other personality scales. However, the most frequent clinical syndromes are generalized anxiety disorders (check Generalized Anxiety scale) and depression, particularly of an agitated nature (check Persistent Depression and, possibly, Major Depression). It is important to consider somatic symptoms as well, as these persons' emotional constriction can lead to physical expression of emotional problems (check Somatic Symptom). Compared to other personality disorders, compulsives tend to be a better-defined population as there is less overlap with other personality disorders. Nonetheless, associated elevations can occur with AASchd/Schizoid (see section on AASchd/Schizoid) and DADepn/Dependent (see section on DADepn/Dependent). Although less frequent, elevations with Compulsive and Narcissistic suggest individuals who are confident, defensive, and unlikely to concede that they have made a mistake. These individuals rely strongly on their own ideas and are likely to have difficulty accepting the advice,

suggestions, and especially the orders of others. Individuals perceive them as inflexible, formal, proper, and distant. As a result, they have difficulty working in supportive team environments where mutual respect and consensus building are crucial factors.

Treatment Implications

Usually compulsive personalities lead controlled, predictable, and generally functional lives. However, when confronted with excessive change or important decisions, they may present to therapy with anxiety-related problems. In particular, these problems might be expressed in somatic complaints because compulsives have a difficult time releasing internal tension. They often view their world in a rigid, inflexible manner. As a result, self-exploration is difficult because it is experienced as a violation of their “character armor,” their personal sense of privacy, and their conformity. In addition, self-exploration runs the risk of playing into their obsessiveness, so that change never actually occurs. One technique for breaking up their obsessive patterns is to help them access and experience their affect. Other strategies are to work with them to realize the irrationality of their patterns or to use paradoxical interventions (e.g., reframing perfection in order to actually allow themselves to make mistakes). Usually the first line of intervention is support, combined with techniques of anxiety reduction: systematic desensitization, relaxation, emotional support, biofeedback, self-hypnosis, and, possibly, psychopharmacological agents. Any insight-related work should proceed cautiously and with considerable reassurance, so that their defenses are not challenged too quickly. Potentially problematic client–therapist transactions might be therapist boredom, power struggles, or therapist collusion with the client’s compulsions in the form of endless but unproductive insights. Despite these potential difficulties, compulsive individuals’ prognosis for treatment is quite good.

Scale 8A: Discontented-Resentful-Negativistic (DRNegat)

The core characteristics measured by this scale are related to an unsettled versus secure and consistent state of thinking and feeling, with resultant inconsistency in attitudes and unpredictability in behavior. These shifts in attitude and behavior are uncomfortable, as no solution is deemed good enough or without regret. The behaviors reflected in this scale are often quite contradictory and erratic.

The *Embittered Expressive Emotion* facet scale reflects the relative ease and frequency that individuals display resentful and bitter feelings and behaviors. Individuals who score high on this facet scale resent others’ contentment and expectations, and they behave in ways to undermine them. They procrastinate and dawdle, stubbornly resisting efficiency and timeliness. They will behave in ways that are contrary and irritating to others, though not outright aggressive or confrontational. They revel in the undermining of others’ accomplishments and happiness. Feedback to such individuals may focus on the resentment they feel when others succeed, as well as the behaviors that arise from that bitterness that undermine others’ success.

The *Discontented Self-Image* facet scale reflects the degree to which individuals feel they have been unappreciated, misunderstood, and generally cheated by fate in life. Individuals who score high on this facet scale continuously feel conflicting feelings of

guilt and resentment, believing that things never go well or right for them, and this is attributable both to their own unworthiness and poor personal qualities and to others' obstruction and own achievements. They are envious of others' achievements but believe success will not happen for them themselves. Things do not work out well for them in life, and they are consequently disillusioned with life and generally pessimistic. Feedback to such individuals may focus on how they define who they are based on failures, mistakes, and missed opportunities, as well as the resentment and bitterness in life that this self-definition breeds.

The *Irritable Mood/Temperament* facet scale reflects the degree of vacillation in mood and general irritability. Individuals who score high on this facet scale are easily irritated and frustrated, erratic in their feelings, and easily driven to negative emotions like anger, upset, and guilt. They have periods of joy and contentment, but these quickly and easily give way to periods of spite, contentiousness, jealousy, and sullenness. They do not hide their emotions well, and they are highly sensitive to even minor slights (or what they perceive to be minor slights). Others' successes elicit envy, jealousy, and ultimately anger that drives oppositional behavior. They may spend quite a bit of time in moody withdrawal, not actively engaging the others who are eliciting negative feelings. Feedback to such individuals may focus on how easily annoyed they are at others, as well as on how sensitive and easily hurt and angered they can be.

Discontented

Individuals who have one of their highest three elevations on the DRNegat but not significant elevation ($BR < 75$) have traits that are considered discontented. These individuals feel that in general the world has treated them unfairly, and they are resentful toward it and others. They are pessimistic and do not expect positive situations to last for long. They are irritable and generally vacillate between pushing forward in their lives and relationships steadily and resentfully withdrawing and undermining themselves and others. Although not necessarily wildly so, these individuals are erratic in their feelings and behaviors. Because they are so erratic, it is difficult for them to act on their environment or others in ways that consistently better their situation. Some withdraw socially in order to avoid the "inevitable" disappointments of life altogether. Feedback to these individuals may focus on their bitter and pessimistic outlook on life, as well as on the behavioral consequences of that outlook.

Resentful

Individuals who have one of their highest three elevations on DRNegat with a moderate elevation have traits that are considered resentful. These individuals are noticeably erratic in their behavior, wavering between resentful oppositionality and resentful acquiescence. Their interpersonal interactions are similarly characterized by often seemingly sudden fluctuation between easy supportiveness and embittered contempt. The reversals of their moods, and consequently of their behaviors, are rapid and frequent. Indirect oppositional behaviors are a frequent result of others expecting anything from them. Feedback to these individuals may focus on their bitterness and

resentment toward others, which bubbles up to the surface quite easily, quickly, and unexpectedly.

Negativistic

The core characteristic for clients with elevations on this scale is a mix of passive compliance combined with resentment and opposition. These clients usually act on these resentments in impulsive and erratic ways. Feeding their resentment is a sense that they have somehow gotten a raw deal in life and will inevitably be disappointed in relationships. However, they also feel that their resentment and anger are not acceptable emotions for them to have. As a result, guilt and conflict pervade their lives. This internal conflictual style also becomes externalized and creates problems in interpersonal relationships. They are moody, complaining, and intermittently hostile. One moment they might be angry and stubborn, but the next moment they feel guilty and apologetic. They are likely to express their negativism in indirect ways—procrastination, inefficiency, and contrary behavior—that have the effect of undermining the happiness of others. They may also act on their resentment with caustic comments, complaints, and expressions of contempt toward others. One means of coping with these feelings is to deny them and instead attribute them to others. Another way is to conceptualize that the resentment and anger are justified because of the numerous reasons to be envious toward others, who are constantly seen as having things so much better. Their resulting chronic unhappiness is expressed through pessimism, disillusionment, and cynicism. Because they blame other people for their misfortunes, they have little insight into how their own behavior and attitudes cause others to reject them. However, when their attitudes and behaviors eventually lead to rejection by others, these clients feel demeaned, abandoned, unappreciated, and disillusioned. Thus, their difficulties are self-fulfilling and self-maintaining. They typically describe themselves as moody, testy, resentful, oppositional, and discontented.

A further core conflict is a feeling that they would like to depend on others, but this dependence is neither socially acceptable nor safe because others inevitably exploit and disappoint them. Thus, they seem moody and unpredictable as they ruminate over these contradictory feelings. They often perceive relationships as a threat to their safety. To protect themselves, they become superficially quite self-sufficient and independent.

At their best, persons with this elevation can be agreeable and friendly. They can also be flexible, changeable, emotionally responsive, and sensitive.

Frequent Code Types

Persons with this code type experience frequent rejection and are likely to experience depression (check Persistent Depression and Major Depression). Their feeling that interpersonal situations are potentially dangerous is capable of producing chronic anxiety (check Generalized Anxiety), which might be expressed in indirect ways through psychophysiological disorders or conversions (check Somatic Symptom). Concurrent elevations on personality scales most often include MPParaph/Paranoid, UBCycloph/Borderline, and ADSadis/Sadistic (see section on ADSadis/Sadistic). Less frequent concurrent elevations include DFMelan/Melancholic, SRAvoid/Avoidant, and AAMasoc/Masochistic. For each of these, the negativistic qualities heighten the

level of anger and resentment, as well as sensitivity to rejection, which can trigger the behaviors consistent with the other personality scales. For example, the concurrent elevation of the Negativistic and Paranoid scales heightens the sensitivity to rejection that fuels paranoid ideation. A concurrently elevated Borderline scale heightens the resentment and sensitivity that triggers disorganization in thoughts, feelings, and sense of self.

Treatment Implications

The two major areas of intervention involve enabling negativistic individuals to be more consistent in their approach to life and to develop insight into the nature of their ambivalent style of responding. However, the therapeutic relationship itself is likely to be complicated by clients' ambivalence. Specifically, they desire caring and support by others but perceive the development of such a relationship as a threat to their independence and fear that it will end up with rejection and disappointment. As a result, they may erratically criticize their therapist or engage in passive resistance. Dealing with this potential difficulty through early behavioral contracting might be particularly useful in keeping these clients engaged in the therapy process. One concern related to clinical management is that their impulsiveness might involve suicide risk. This fact is especially problematic if they decompensate into an anxiety or depressive disorder. Family and marital interventions are likely to be extremely beneficial because negativistic patterns are both initiated by and maintained in these systems. Formal programs of anger management and assertiveness training (e.g., dialectical behavior therapy) might also be quite helpful in helping these clients to develop greater control over impulses and learning more effective styles of communication. Their belief in future disappointments, along with their dysfunctional thoughts of having been cheated by life, can be worked on through cognitive interventions that challenge these assumptions. Because they are likely to be resistant, controlling clients, the use of either paradoxical directives or a combination of nondirective and client-directed techniques is likely to optimize outcomes.

Scale 8B: Abused-Aggrrieved-Masochistic (AAMasoc)

The core characteristics measured by this scale are related to the experience of pain and hurt as somehow preferable to positive emotions. The self-defeating nature of these characteristics often represent having some control over situations (controlled failure is often preferable to earnest trying with the chance for genuine failure), and the negative emotional states are often mild in nature. That is, purposeful, mild negative emotional states may be preferable to the potential for extremely negative emotional states that may occur outside of an individual's control. It should be noted that it has been argued that the characteristics measured by this scale have an antifemale sex bias (Kutchins & Kirk, 2003), though others have argued that sex bias results in inappropriate use of the characteristics rather than the category itself (Fuller & Blashfield, 1989). Clinicians should consider the issue of sex bias and gender-biased social norms carefully when interpreting results on this scale.

The *Undeserving Self-Image* facet scale reflects the degree to which individuals feel they deserve the worst in life because they themselves are not good enough. Individuals

who score high on this facet scale focus on their weaknesses, shortcomings, and overall worst personal qualities. They undervalue their opinions, and as a result they quickly give in when others disagree with them. They do not feel they deserve kind gestures or to do things for themselves, keeping their expectations and wants to a minimum. They feel they have not and will not live up to others' expectations and that they deserve shame, humiliation, and negative general outcomes. Feedback to such individuals may focus on their overwhelmingly negative view of themselves and the belief that they deserve negative consequences because they are not good enough.

The *Inverted Intrapsychic Architecture* facet scale reflects a switching of internal experience between pain and pleasure. Individuals who score high on this facet scale have a component of their inner world that experiences pain when pleasure is more normatively appropriate and pleasure when pain is more normatively appropriate. It is important to note that this is not the entirety of their experience. Much of their internal world is conventionally built, with drives and urges to meet basic needs for pleasure and satisfaction. However, the part of their internal experience that is mixed drives these individuals to behave in ways that appear the opposite of what most people would want, as they try to elicit negative consequences. Their internal experience suggests to them that in self-defeat, not only will they avoid others controlling the situation in a way that hurts or humiliates them, but they will also somehow elicit nurturance. Feedback to such individuals may focus on the high level of comfort and satisfaction they feel with situations and consequences that would make most other people uncomfortable or upset, as well as the occasional (if not often) discomfort with entirely positive experiences.

The *Dysphoric Mood/Temperament* facet scale reflects the degree of complexity and generally negative emotions felt by individuals. Individuals who score high on this facet scale suffer emotionally. Their emotional experience is a complex array of mixed anxiety, anguish, torment, hopelessness, guilt, and misery. Their complex, negative typical emotional state is generally quite salient in their demeanor and affect, which elicits guilt, helplessness, and often ultimately annoyance in others. The internal emotional world of these individuals is broadly dysphoric. Feedback to such individuals may focus on the broad array of negative emotions they feel at any given moment.

Abused

Individuals who have one of their highest three elevations on the AAMasoc but not significant elevation ($BR < 75$) have traits that are considered abused. These individuals are self-sacrificing and most often delay meeting their own needs until the needs of others have been fully attended to. They are selfless and feel they deserve love and affection only when they are being helpful to others and serving to meet others' needs. They are often deferential and obsequious, feeling undeserving of attention and unsure of their own opinions. They are not particularly comfortable with themselves, and as a result they focus their lives on other people and their needs and wants. Some "abused" individuals have very little faith in the lasting nature of relationships and as a result do not engage freely and easily in them. This subgroup of abused individuals tends to be relatively self-sufficient, though when forced to interact, these individuals

will do so in a yielding, subservient manner, in order to avoid conflict or draw undue attention to themselves. Feedback to “abused” individuals may focus on their general self-sacrificing and often altruistic stance toward interpersonal relationships.

Aggrieved

Individuals who have one of their highest three elevations on AAMasoc with a moderate elevation have traits that are considered aggrieved. These individuals are self-sacrificing and focused on the needs of others before their own, to the point that it deteriorates their self-worth and reinforces a generally negative emotional experience. They display more overtly self-destructive and self-sabotaging behaviors, often within relationships with other people. Some of these aggrieved individuals take pride in their self-sacrifice and other focus, finding both a sense of identity and self-esteem in denying their own needs and serving others. They expect consistent displays of appreciation from others, and they react quite negatively (both emotionally and behaviorally) when they feel others are underappreciating their sacrifices and loyalty. Some of these aggrieved individuals work hard to make themselves necessary in the lives of others, eliciting guilt from others who try to separate from them, even temporarily. These individuals are possessive in their self-sacrificing ways, often intrusive in placing themselves in situations as helpers or advice givers. Feedback to aggrieved individuals may focus on their extreme degree of self-sacrifice and the consequences of their deprecating behaviors, both for others and for themselves.

Masochistic

High elevations on Masochistic indicate aggrieved persons who continually place themselves in situations in which they will be the victims. They present themselves as inferior, nonindulgent, self-effacing, insecure, or otherwise reluctant to accept pleasure and happiness. Somehow, pleasure is seen as something they do not deserve, and they feel that if they allow themselves to experience pleasure, further difficulties or other unpleasant consequences will follow. Anything positive is expressed with very little enthusiasm. These clients characterize interpersonal relationships as servile, self-effacing, self-sacrificing, or otherwise allowing or even encouraging others to exploit or mistreat them. This active involvement in creating situations in which they will be exploited differentiates these types of persons from other depressed clients. Close relationships are usually associated with disappointments and frustrations. Those who do try to support and help them are likely to be ignored or otherwise rendered ineffectual. One purpose of this response is for clients to make themselves appear weak and harmless in an effort to discourage possible criticism and aggression from others and evoke guilt instead. In addition, their public displays of dejection initially produce both sympathy and a tacit permission to avoid unpleasant responsibilities. A further purpose is to keep their self-identity organized around being shamed, humbled, and debased. They may be so absorbed in their own suffering and misery that they have few resources for appreciating the dilemmas others might be in. Although they might be superficially sympathetic to others, underneath they are unempathic and distrustful. They focus and ruminate on past failed relationships and

disparage any personal achievements. This results in their being anxious, apprehensive, mournful, anguished, and tormented.

Positive qualities are that, in comparison to disorders such as schizoid, these clients are involved with and connected to people. They can often develop a good level of insight into their difficulties. In addition, their level of distress is likely to be sufficiently high that they can and do become engaged in therapy.

Frequent Code Types

The greatest risk for masochistic persons is the development of depression (check Persistent Depression and Major Depression). If anxiety is present, it is usually diffuse and associated with fears of loss and abandonment, although Generalized Anxiety may be elevated. Hypochondriacal strategies might be grafted onto their aggrieved style as a means of channeling anxiety and obtaining support (check Somatic Symptom). The most frequent associated elevations are with DADepn/Dependent (see section on DADepn/Dependent), UBCycloph/Borderline (see section on UBCycloph/Borderline), DFMelan/Melancholic (see section on DFMelan/Melancholic), and SRAvoid/Avoidant. When Masochistic and Avoidant are both elevated, it suggests that these persons have found relationships sufficiently painful that they have withdrawn to the extent of rarely interacting and becoming relatively isolated. They would like to be involved with others, but that experience has simply proven to be too painful in the past.

Treatment Implications

The paradox of working with masochistic persons is that the context of therapy is to make them happier; yet, on one level, they do not want to be happier. These clients might even try to provoke or at least frame situations in such a way that they feel rejected or humiliated by the therapist. To counter this tactic, a sufficient amount of support, understanding, and rapport must be established to work with these clients and make them understand that they do not necessarily have to suffer. Specific self-defeating behaviors need to be identified along with the circumstances that elicit them. Assertiveness training, to help clarify their rights and develop skills to stop exploitation, might be particularly helpful. These skills, and others, might be practiced in the context of role plays and/or couples therapy. Further examination of relationships and the part clients play in them can occur both in individual therapy and through supportive group interaction.

SEVERE PERSONALITY PATHOLOGY

The three severe personality pathology scales work slightly differently from the other personality scales. Millon et al. (2015) set these up, based on Millon's (2011) work, such that the middle level of each is quite seriously pathological. That is, along these spectra, there is not the same normal–slightly abnormal–abnormal continuum but rather a normal–abnormal–significantly abnormal continuum, as there is less chance of healthy functioning along these spectra. Clinicians should familiarize themselves with the MCMI-IV's terminology for these scales in order to better interpret them.

Scale S: Eccentric-Schizotypal-Schizophrenic (ESSchizoph)

The core characteristics measured by this scale are related to eccentricity, eschewal of social convention, seemingly inappropriate emotional expression, and oddity and bizarreness in thinking and behavior. These characteristics in their most severe form represent a complete psychotic break from reality, along with the thought, emotional, and behavioral correlates that accompany such a break.

The *Circumstantial Cognitive Style* facet scale reflects the degree of difficulty individuals have in organizing their thoughts. Individuals who score high on this facet scale perceive and interpret the world and its cues in an idiosyncratic, normatively unusual way. They have difficulty correctly interpreting social cues, attributing their own personalized and often bizarre meaning to them. As a result, they often distort the meanings of social interactions. Even relatively normal or boring social interactions are imbued with ideas of reference, odd metaphorical meaning, and circumstantial logic. As a result of the idiosyncratic way they perceive and interpret the world, they often develop magical thinking, suspicious thoughts, bizarre logic, and difficulty differentiating between reality and fantasy. Their bizarre thoughts, rooted in their unusual thinking process, often occupy their minds in a way that becomes ruminative. Feedback to such individuals may focus on unusual meaning that they ascribe to even normal situations, as well as clear indication that their thinking processes do not follow the same method as most other people's.

The *Estranged Self-Image* facet scale reflects feelings that individuals are alienated from others and dissociated from the world in some way. Individuals who score high on this facet scale feel estranged from other people and the world around them, often contemplating the meaninglessness of life. They often feel dissociated from the world, experiencing depersonalization (having the experience as if they are observing themselves from outside their body) and derealization (having the experience that they are somehow not real and rooted in reality). These individuals often view themselves as being glum, focusing on their purposelessness in life. Feedback to such individuals may focus on their experiences of feeling unconnected to the real world or other people.

The *Chaotic Intrapsychic Content* facet scale reflects chaos in the internal world of individuals. Individuals who score high on this facet scale have internal templates for interacting with the world that are jumbled, chaotic, and random. Without a coherent, consistent template for what to expect when interacting with others and the world in general, these individuals become overwhelmed by random impulses, lost in tangential thoughts, and disoriented by scattered expectations. The higher the score on this facet scale, the more chaotic and confusing the internal world of the person is. Feedback to such individuals may focus on just how overwhelming and confusing their thoughts and feelings can be.

Eccentric

Individuals who have one of their highest three elevations on the ESSchizoph but not significant elevation ($BR < 75$) have traits that are considered eccentric. These individuals are not so odd and withdrawn that they cannot function effectively in the everyday world. However, these individuals do behave in peculiar ways at times. They are often notably shy, but when they do interact they at times appear odd and unusual.

Their emotional expression at times seems disconnected and even contradictory to what they are saying. Although not necessarily aware of their bizarreness or eccentricity, these individuals often do have a feeling that they are somehow different and separate from most other people. This may make them somewhat disengaged or apprehensive in their interpersonal interactions, which may appear to others as withdrawn, shy, and even oddly self-sufficient. Feedback to such individuals may focus on the fact that they do not feel constrained by social norms of appropriate behavior and that they may be slightly wary of interacting with others because of their feelings of separateness and difference from them.

Schizotypal

Individuals who have one of their highest three elevations on ESSchizoph with a moderate elevation have traits that are considered schizotypal. These individuals think and behave in ways that increasingly lose touch with reality. Their thoughts and feelings are haphazard and tangential, and they keep a great deal about themselves a secret. They often experience depersonalization and derealization, experiencing their thoughts and feelings as separate from their body. Some actually feel “dead” and lifeless, lacking vitality and engagement in the world. Most schizotypal individuals appear sluggish and affectless. Many withdraw significantly from the social world, preferring to live primarily within their fantasy worlds. They have little hope for interpersonal relationships, and their odd and disorganized thinking and behavior is not conducive to having them. Feedback to such individuals may focus on their disorganized and confusing thoughts, as well as on their tendency to be secretive and withdrawn from others.

Schizophrenic

The major characteristics of persons with schizophrenic-level elevations on are eccentricity, disorganization, and social isolation. Although the term *schizophrenic* is used, characterizing the difficulties as similar to the more episodic clinical syndrome, the psychotic difficulties reflected on this scale are usually of a long-term nature. The eccentricities of these persons relate to peculiar mannerisms, strange clothes, and bizarre expressions. They typically look drab, lifeless, apathetic, and joyless. Self-descriptions include alienated, isolated, fragmented, and detached. They may engage in magical behavior and rituals in an attempt to neutralize “evil” thoughts, deeds, or omens. Often there is little distinction between fantasy and reality. Their communication style is characterized by tangential comments, personal irrelevancies, and magical associations. As a result, they lead empty and personally meaningless lives in which they drift to and from various locations and sources of employment. Thus, they exist on the fringes of society. Some are detached and emotionally bland; others are more suspicious, anxious, and apprehensive. Because they are mistrustful and communicate poorly, their relationships usually make them quite uncomfortable. As a result, they develop few, if any, close friendships and prefer privacy, secrecy, and isolation. Usually they lack the interest and energy to initiate social interaction. Internally, they have a deep sense of emptiness and meaninglessness, which is sometimes sufficiently severe to prompt a full schizophrenic episode. Their thought processes are scattered, autistic, and disorganized. They are likely to have experiences of depersonalization and dissociation.

These individuals are cognitively impaired in their ability to comprehend interpersonal motivations and communications.

Frequent Code Types

Diagnostically, schizotypal personalities exist somewhere between the less severe schizoid disorder and the more severe schizophrenic disorders. However, there is conceptual and clinical overlap with both these disorders; therefore, elevations on scales that measure these dimensions should be noted (check AASchd/Schizoid, Schizophrenic Spectrum, and Delusional). Accordingly, schizoid and schizophrenic disorders might coexist with schizotypal. The most likely associated elevations on personality scales are AASchd/Schizoid, SRAvoid/Avoidant, and MPParaph/Paranoid. The Schizoid and Avoidant elevations are important in distinguishing two subtypes of schizotypal personalities. An elevation on Schizotypal in combination with Schizoid indicates a more passive, apathetic, detached expression of schizotypal characteristics. These persons are deficient in their capacity to experience emotions and extremely detached and indifferent toward others. In contrast, an associated elevation on Avoidant indicates a desire for personal contact, but these individuals are more anxious and apprehensive and actively protect themselves by disengaging from others. If Paranoid is elevated along with Schizotypal, it highlights these clients' suspiciousness, along with corresponding ideas and delusions of reference. Although their thoughts might be more organized because of the coherence provided by the paranoid content, they still have the tangential thinking and eccentric behavior that are characteristic of persons with elevations on Schizotypal.

Treatment Implications

The prognosis for schizotypal personalities is not good because of the ingrained, long-standing nature of their patterns and the difficulty of engaging them in the therapeutic process. Treatment goals should be tempered accordingly, with a focus on preventing further social isolation and deterioration. Changing these individuals' environment to encourage an increase in supportive interpersonal interaction might be particularly helpful. A further intervention might be to help them express and clarify their thoughts while simultaneously providing emotional support. Psychopharmacological agents might be useful both in helping to organize their thoughts and in reducing the likelihood of their acting on irrational impulses.

Scale C: Unstable-Borderline-Cyclophrenic (UBCycloph)

The core characteristics measured by this scale are related to internal conflict, flux, and ambivalence experienced by individuals. Additional characteristics measured by this scale relate to unpredictable behavior and general inconsistency in thoughts and feelings about self and others.

The *Uncertain Self-Image* facet scale reflects the degree to which individuals do not have a core sense of who they are and an understanding of their own identity. Individuals who score high on this facet scale do not have a stable, unwavering sense of who they are as people. They are confused about their identity and have difficulty

settling on a single endeavor that might help them define themselves more clearly. Their self-presentation shifts, either from situation to situation or even within a single situation, and these presentations of themselves are often contradictory. Most often these individuals feel a sense of emptiness underneath their fluctuating sense of who they are. Feedback to such individuals may focus on their lack of consistency in who they are from one situation or moment to the next.

The *Split Intrapsychic Architecture* facet scale reflects the degree of segmentation and conflict that exists within the inner world of a person. Individuals who score high on this facet scale have conflicting and highly opposed inner “maps” to guide them in their interactions with others and the world. For example, some of their internal workings focus on longing for closeness, nurturance, and support of others in order to feel secure and safe. Other internal workings, however, focus on the inevitable pain, disappointment, and abandonment they expect from any significant relationship. These opposing forces within cause significant turmoil, not providing any clear sense of how they should proceed in interpersonal interactions and relationships. Their internal world provides them with a multitude of no-win situations when they interact with the world. Feedback to such individuals may focus on the deep ambivalence (severely mixed feelings) they tend to feel, especially about interpersonal relationships.

The *Labile Mood/Temperament* facet scale reflects the degree to which emotions frequently and dramatically shift in an unmodulated and uncontrolled manner. Individuals who score high on this facet scale have intense and rapidly shifting emotional states that can easily confuse, upset, and disorient them. Often the emotional states of these individuals are not aligned with reality and situations outside of their minds. These individuals quickly and dramatically shift between normal, situation-based moods to intense and inappropriate anger, with frequent periods of dysphoria and even some anxiety or excitement. These unrestrained and fluctuating emotional states have a strong influence over their behavior. Feedback to such individuals may focus on their dramatic and unpredictable shifts in mood.

Unstable

Individuals who have one of their highest three elevations on the UBCycloph but not significant elevation ($BR < 75$) have traits that are considered unstable. These individuals are not so impulsive and unpredictable that they cannot function effectively in the everyday world. However, they are highly emotionally reactive in interpersonal relationships in ways that often do not seem logically related to the situation at hand. They most often have intense and rocky relationships, owing either to their frantic and manipulative tactics for ensuring that others engage with them or to their sabotaging relationships so that they do not feel engulfed by them. Both of these patterns stem from emptiness and insecurity about their own identity, either desperately needing the connection to another to define them or fearing that they will somehow lose themselves and be absorbed if they commit too deeply to a relationship. These individuals tend to have a fragile sense of self-worth, tied both to their unstable self-image and to their unstable pattern of relationships. Feedback to such individuals may focus on their intense emotional reactions within interpersonal relationships, which have led to (and been the result of) a chaotic social life.

Borderline

Individuals who have one of their highest three elevations on UBCycloph with a moderate elevation have traits that are considered borderline. These individuals exhibit significantly labile mood, thoughts, and behaviors. Their expressed emotions are often inappropriate to the situation, more dependent on internal turmoil than on reality. Their erratic behavior causes interpersonal difficulties. Their ambivalence about closeness with others is marked, with quite strong dependency needs as well as quite strong expectations for being abandoned. Their thoughts are often heavily influenced by their labile mood, resulting in suspiciousness and intense anger. They are irresponsible and impulsive, driven by their intense feelings. Some seem restless, impatient, and irritable, and brief psychotic episodes are not uncommon. Feedback to such individuals may focus on their tumultuous lives, with intense emotions, chaotic relationships, and unpredictable and impulsive behaviors.

Cyclophrenic

The core features of individuals with elevations on this scale are instability and unpredictability of mood and behavior. One moment they might feel dejected and disillusioned; sometime later, feelings of euphoria are followed by a phase of intense anger, irritability, and self-destructiveness—possibly even involving self-mutilation. Their self-destructiveness reflects a severely punishing conscience. In addition, much of their unstable behavior seems to be directed by internal factors rather than a reaction to environmental events. They have marked mood swings, intermittent periods of depression, generalized anxiety, and intense emotional attacks on others, followed by apathy and dejection. Although these behaviors often create significant interpersonal difficulties, these individuals are also extremely concerned with maintaining the care and emotional support of others. They often elicit rejection, but they react strongly to fears of abandonment. They might intermittently idealize people, but their ambivalence eventually gives way to devaluing and criticizing the same people they have previously idealized. Thus, their relationships are characterized by ambivalence, instability, and intensity. Underlying many of these behaviors is an extremely poorly developed sense of identity, which is at the core of their dissolution of controls. Their poorly defined sense of self might eventually give way to feelings of emptiness and to disorganized thoughts. Under stress, they often have transient psychotic episodes. However, these episodes are rarely sufficient to be considered a formal thought disorder, and these clients usually return fairly quickly to their previous levels of functioning. They typically describe themselves as depressed, impatient, tense, irritable, disturbed, and anxious.

Frequent Code Types

The symptomatology of borderline personalities can be extremely diverse; elevations may appear on any of the clinical scales. However, mood disorders (check Bipolar Spectrum, Persistent Depression, and Major Depression) and substance abuse (check Alcohol Use and Drug Use) are among the most common complications. In many ways, borderline personalities can be conceptualized as exaggerations or extensions of the less dysfunctional personality disorders of masochistic, negativistic, dependent,

histrionic, and/or narcissistic. As a result, elevations on one or more of the scales representing these constructs would be expected and would provide further information on these individuals' underlying dynamics and particular mode of expression. Because the borderline category encompasses such a broad spectrum of behaviors, integrating the meaning of other scale elevations can be crucial information to attend to. One of the most frequent associated scale elevations is when Borderline is combined with DRNegat/Negativistic, which emphasizes the conflicted aspect of the borderline personality. These clients feel intense dependency yet are anxious and extremely ambivalent about it. They also feel intense resentment and anger but simultaneously believe that such feelings are unacceptable. These intense polarities might naturally give way to both a disintegration of the sense of self and clearly unstable, unpredictable behavior. Another important combination is Borderline and AAMasoc/Masochistic, which would highlight these clients' impulsive and self-destructive characteristics. Behind their unstable emotions and behavior lies a strong underlying sense that they are not worthy of happiness but instead should be exploited and humiliated. Thus, dealing with the client's depression and potential suicidality would be an essential aspect of case management. Elevations on DADepn/Dependent and Borderline emphasize these clients' low self-esteem, passivity, and apathy, combined with their need for someone else who will care for them and make decisions for them. A corresponding elevation on SPHistr/Histrionic would underscore these persons' dependency, but, instead of being apathetic and passive, they would be outgoing, friendly, manipulative, and emotional. When their defenses are challenged, they might increase their activity and attention seeking to intense levels; if this strategy does not work, they may deteriorate into futility and self-destructiveness. When CENarc/Narcissistic is elevated along with Borderline, it suggests that these individuals' self-inflated sense of importance has collapsed into feelings of shame, insecurity, emptiness, and self-condemnation.

Treatment Implications

Although borderline personalities are notoriously difficult to work with, they are also more amenable to change than many other individuals with personality disorders. The central, initial goal is to build sufficient rapport so that work can begin on stabilizing their erratic behavior and affect. Doing this might involve a reality-oriented approach emphasizing aspects such as limit setting, sympathy, reassurance, advice, and insight regarding internal processes. Borderline personalities are capable of a wide range of dysfunctional behaviors, but the highest priority should be given to working with suicidal and self-injurious behavior. In addition, they are an unusually heterogeneous group. For example, depression, anxiety, depersonalization, disorganized thoughts, fears of abandonment, self-destructiveness, and/or ambivalence may all become areas requiring attention. More than for most other client groups, building a strong therapeutic alliance is crucial in helping borderline personalities to adjust and cope with their many conflicted forms of acting and feeling. Effective treatment has often been achieved using a combination of mindfulness skills, techniques to assist with regulating emotions, strategies to assist with interpersonal effectiveness, and building tolerance to distress (such as in dialectical behavior therapy). Because many borderline personalities resist authority-directed interventions, group therapy might be indicated, as they are more likely to be responsive to peer influence.

Scale P: Mistrustful-Paranoid-Paraphrenic (MPParaph)

The core characteristics measured by this scale are related to constriction and unyielding inflexibility (even despite external evidence) in the way individuals view the world, specifically reflecting suspiciousness, expectation for deception, and resultant angry avoidance of closeness with others. The rigid way individuals high on this scale experience the world around them often results in hostility and aggressiveness toward others.

The *Defensive Expressive Emotion* facet scale reflects the degree to which individuals are guarded with their expression of emotion. Individuals who score high on this facet scale are highly influenced by their mistrustful cognitive style (see the next section titled “Mistrustful”), which renders them hyperalert to potential danger and in turn tense and constricted in their emotional expression. These individuals are often irritable, even with minor provocation, and quickly and easily become hostile. In general, though, their demeanor is one of tension, constriction, and defensiveness. Feedback to such individuals may focus on their level of tension and how quickly and easily they become irritated by, annoyed by, or angry at others.

The *Mistrustful Cognitive Style* facet scale reflects the hallmark of the pathology described by the overall scale, the degree of suspiciousness and cynicism toward others. Individuals who score high on this facet scale are markedly mistrustful of the behaviors of others, often attributing negative intentions and motivations to even benign or ambiguous interactions. They tend to highlight and catastrophize minor or irrelevant details that confirm their suspicions of negative intentionality on the part of others, and they are constantly on the lookout for deception and malice. These individuals are constantly attending to any aspect of situations that may confirm their suspicions, and the higher they score on this facet scale, the more likely they are to distort information to supply evidence or, in the absence of any evidence, to attribute their lack of confirmation to the ultimate cunning and deceptiveness of others. Feedback to such individuals may focus on their high baseline level of suspiciousness and their expectation that others are ill-intentioned in general.

The *Projection Intrapsychic Dynamics* facet scale reflects a mechanism of disowning their own deficiencies and negative characteristics and attributing them to others. Individuals who score high on this facet scale are unaware of their own negative attributes, including their own hostility toward others. All of their negative or intolerable qualities are projected onto (attributed to) others in their life, even with little to no evidence. That is, even insignificant, inconsequential, and at times nonexistent negative qualities in others are salient to these individuals. Without being aware of their own deficiencies, these individuals can easily build up an image of themselves as superior to others, as well as the innocent victim of others’ maliciousness. Higher scores on this facet scale reflect a higher likelihood that their projections have led increasingly to a view of themselves and others that is not based in reality, but rather in fantasy. Feedback to such individuals may focus on their lack of awareness of their own faults and negative qualities, despite their keen awareness of these traits in others.

Mistrustful

Individuals who have one of their highest three elevations on the MPParaph but not significant elevation ($BR < 75$) have traits that are considered mistrustful.

These individuals are not so suspicious and oppositional that they cannot function effectively in the everyday world. However, these individuals are vigilant and highly aware of the potentially malicious behaviors of others. Uncomfortable relying on others for anything, these individuals tend to be entirely self-sufficient and even controlling of others in their lives. They can often be self-righteous, disgusted by weakness and faults in others and viewing themselves as invulnerable and superior. Some of these individuals are quite rigid and controlled in the way they approach the world, such that they are rule-bound and stubborn, tense, and humorless. This rigidity is an effort to clarify and capitalize on a path toward power and autonomy, so they do not have to rely on others. Feedback to such individuals may focus on their self-sufficiency and the fact that it stems from a general underlying sense that they cannot trust others to support them, get things done appropriately and accurately, and put aside their own self-interests in order to help the mistrustful individual.

Paranoid

Individuals who have one of their highest three elevations on MPParaph with a moderate elevation have traits that are considered paranoid. These individuals have overtly paranoid symptoms, with marked suspiciousness and idiosyncratic and most often non-reality-based beliefs about the malicious and cunning behavior of others. They accuse others of negative motivations and intentions, aligning “evidence” in their minds of apparent and obvious malice. This evidence is most often tangential, loose, and fabricated, by others’ perception, but it makes clear sense to the paranoid individuals and confirms their suspicions. These individuals often exhibit clear delusions, separate from reality. Some are driven by these delusions to be hostile, oppositional, and irritable. These individuals feel misunderstood, put upon, and actively targeted and harmed by others. Other paranoid individuals are driven by their delusions to withdraw from the world significantly and isolate themselves. These individuals feel highly vulnerable and sensitive to the “inevitably” harmful behaviors of others. They are markedly hypervigilant, ever alert to the possibility of harm in a destructive and hostile world. Feedback to paranoid individuals may focus on their suspicious outlooks on life, including specific beliefs about others being out to harm them, and the resultant behaviors (hostile or isolating, which can be better understood by which other scales are elevated along with this one).

Paraphrenic

The central issue for persons with elevations on MPParaph is suspiciousness and defensiveness, combined with a feeling of superiority. They are constantly vigilant because they feel others will criticize or deceive them. Innocuous events are perceived as insults or as the workings of a world in which others are trying to control or harm them. They distort their world by interpreting events to fit their idiosyncratic views. Because they feel in frequent danger, they are abrasive, touchy, hostile, and irritable. They are likely to feel bitter toward people who have been successful and to believe that their success has been achieved through dishonesty and possibly illegal activities. This process involves denying their own shortcomings and attributing them to others. Although quick to notice and expand on minor faults in others, they are ignorant of these same faults in themselves. These dynamics are used as a means of establishing their own

superiority in relation to others. They often describe themselves as misunderstood, righteous, suspicious, mistreated, and defensive.

If these individuals perceive that anyone is trying to control or influence them, they consider this a personal encroachment on their independence and will attack and humiliate the encroacher. As a result, they frequently induce fear and exasperation in others. Unfortunately, their system of making sense of the world is self-fulfilling. People react negatively to the mistrust and even hostility of these individuals, which provides evidence that indeed the world is a dangerous, insecure place. When other people act in negative ways toward them, this pushes them progressively into a more insular world in which their thinking becomes extremely rigid. The rigidity and insularity are maintained because they depend on their own internal processes for both stimulation and reinforcement. They are terrified of being dominated and consider any sign of dependence an indication of weakness and inferiority. They insist on being the designers of their own fate and, to do so, need to be free from entanglements and obligations. Behind this separateness is a fear of losing their personal control and sense of autonomy. Thus, their extremely tightly organized and coherent personality and cognitive structure makes them feel emotionally and physically disconnected from others. These individuals may have delusions of grandeur, delusions of reference, and intense fears of persecutory plots and conspiracies.

Frequent Code Types

Given the mistrust and fear expressed by many paranoids, anxiety is probably the most frequent Axis I complication (check Generalized Anxiety). Additional difficulties are likely to be obsessive-compulsive syndromes in which they engage in compulsive activities in an attempt to make their world “safe.” In severe paranoid states, psychotic symptoms, expressed through delusions and hallucinations, may be present (check Schizophrenic Spectrum and Delusional scales). Related elevations on personality scales most frequently include CENarc/Narcissistic, UBCycloph/Borderline, and SRAvoid/Avoidant. If Narcissistic is elevated along with Paranoid, it suggests that, at some earlier stage, these individuals’ self-inflated sense of importance and superiority has been severely challenged. Paranoid processes become a means to resurrect these beliefs in a way that is further separated from reality and therefore requires more drastic measures. The result might be extravagant plans to defend the world from evil, create new societies, or solve insurmountable scientific problems. Concurrent elevations on Borderline suggest that intense hostility, irritability, and anger are most characteristic of the expression of the underlying paranoid beliefs. These individuals are emotionally and ideationally erratic, though the swings are generally from one negative state to another, rarely to a positive or euphoric state (with the exception of some manic grandiosity). Elevations on Avoidant and Paranoid indicate that these clients are handling their fears and suspicions by becoming progressively more insular, reclusive, and isolated. Insularity helps to protect them from fears that others will be able to influence their thought processes. However, they also feel extremely vulnerable and have serious questions related to their self-esteem.

Treatment Implications

Although paranoid personalities have an intact, organized means of processing their world, they develop and maintain this perspective by insulating themselves from

the influence of others and developing extremely rigid cognitive structures. Because therapy tries both to influence clients and to loosen habitual ways of perceiving the world, these individuals are difficult to work with. As a result, their prognosis is poor. Furthermore, submitting to therapy is an admission of weakness and of giving up self-sufficiency, and both situations are abhorrent and threatening to them. A therapist who is too friendly and empathic is likely to be perceived as being deceitful. High empathy by the therapist has even been found to be counterproductive. In contrast, a therapist who is too distant or who challenges these clients' delusions will seem rejecting. Either approach may, therefore, invoke the clients' suspicions. The relationship requires a delicate balance. Trust needs to be slowly built up with gradual but careful encouragement to perceive events from several different perspectives. Often a goal of treatment is to facilitate and help maintain psychopharmacological interventions and compliance.

CLINICAL SYNDROMES

Scale A: Generalized Anxiety

High scores indicate clients are complaining of tension, difficulty relaxing, indecisiveness, and apprehension. Additional complaints include a highly sensitive startle response, hyperalertness, and fears related to the onset of poorly defined difficulties. Physiological complaints related to overarousal are also common. These might include insomnia, headaches, nausea, cold sweats, upset stomach, palpitations, excessive perspiration, and muscular aches. Anxiety may be either generalized or more focused, as in social situations or specific phobias. Inspection of responses to individual scale items can help to assess the degree of specificity of the anxiety.

Scale H: Somatic Symptom

Elevations reflect somatic complaints expressed in areas such as generalized pain, fatigue, multiple vague complaints, and/or preoccupation with health-related difficulties. However, these often represent psychological conflicts that are being expressed through physical means. If clients have legitimate physical illnesses, they are likely to be unduly preoccupied and possibly exaggerating their difficulties. In other words, they overinterpret their difficulties to signify a major illness when the illness is actually relatively minor. Often the complaints are expressed in a dramatic and/or vague manner. An important function of these complaints is to gain sympathy, attention, or medical reassurance. A careful medical history typically reveals a hypochondriacal pattern in which these individuals are overusers of the health care system. However, when interpreting this scale, it is important to note from clients' history if they are in fact currently suffering from a major medical illness.

Scale N: Bipolar Spectrum

High scorers are likely to have mood swings that range from elation to depression. When elated, they are restless and distractible, have an exaggerated sense of self-esteem, and are overly optimistic and impulsive. They have a heightened and general sense of

enthusiasm along with unrealistic goals. Interpersonal relationships have a demanding, intrusive, and pressured quality. There is a reduced need for sleep, erratic mood shifts, and flighty ideas. Extreme elevations generally indicate a psychotic process characterized by delusions and possibly hallucinations.

Scale D: Persistent Depression

Elevations on Persistent Depression reflect sadness, pessimism, hopelessness, apathy, low self-esteem, and guilt. These persons continuously feel socially awkward, introverted, sad, useless, and filled with self-doubt. Discouragement and a preoccupation with their own inadequacy are also present. They have a sense of futility and may easily break into tears. Somatic complications might include insomnia, a poor appetite or habitual overeating, poor concentration, a continuous sense of feeling tired, and a marked loss of interest in pleasurable activities. Although these individuals may have reduced effectiveness in competently undertaking daily activities, they still remain involved in everyday life. Suicidal ideation might be present and should be investigated further. This, and other details related to the nature of the depression, can be further understood by noting the responses to particular items. Unless the Major Depression scale is markedly elevated as well, it is unlikely that the depression will be sufficiently severe to include psychotic symptoms.

Scale B: Alcohol Use

Individuals scoring high on Alcohol Use are likely to have had a history of problem drinking. They may have tried unsuccessfully to curb or discontinue their drinking. High scorers are also likely to be having social, family, and/or occupational distress. However, the degree to which their drinking is problematic needs to be assessed in relation to other information about their level of functioning.

Scale T: Drug Use

High scorers will have had a recurrent history of difficulties with drug abuse. Also present are a number of traits associated with drug-related difficulties: hedonism, impulsiveness, difficulty conforming to mainstream standards of behavior, self-indulgence, exploitiveness, and narcissistic personality characteristics. High scorers are likely to have difficulty organizing daily life activities and experience social, family, legal, and/or occupational distress.

Scale R: Posttraumatic Stress

Elevations on this scale suggest that these individuals have experienced an intense, life-threatening event that has resulted in extreme fear, helplessness, and arousal. As a result of the event(s), they have uncontrolled, intrusive, and recurrent images or emotions related to the event(s): flashbacks, nightmares, or dissociative feelings that reactivate the event(s). Anxiety-related symptoms might include hypervigilance, hyperalertness, overreactive startle reactions, and a compulsive avoidance of circumstances that might be related to the trauma.

SEVERE SYNDROMES

Scale SS: Schizophrenic Spectrum

High scores on Schizophrenic Spectrum suggest these persons have thoughts that are inconsistent, bizarre, fragmented, and disorganized. In addition, their behavior might be regressed, secretive, and incongruous; and they might be confused, withdrawn, and disoriented. Their affect is likely to be blunted, and they may report hallucinations. Possible diagnoses include brief psychotic, schizophrenia, and schizophreniform disorders, as well as psychotic features related to other disorders, such as mood disorders.

Scale CC: Major Depression

High scores suggest severe depression, to the extent that these individuals have difficulty with effective daily living. Psychological difficulties include a sense of hopelessness, suicidal ideation, pessimism, ruminating, and fear of the future. Somatic symptoms might include insomnia, poor concentration, psychomotor slowing or agitation, loss of appetite, weight loss, chronic fatigue, early-morning awakening, and loss of sexual desire. They are also likely to feel worthless and to experience guilt. Some high scorers might express their symptoms in an irritable, whining manner, whereas others might be shy, passive, seclusive, and introverted.

Scale PP: Delusional

Elevations on this scale indicate acutely paranoid states. These individuals are characterized by irrational but interconnected delusions, persecutory thoughts, and grandiosity. They are hyperalert to possible threats. The most frequent mood is hostile suspiciousness, perhaps to the point of belligerence. They feel mistreated, jealous, and betrayed. They rigidly hold onto their paranoid beliefs despite evidence that they are not true.

RECOMMENDED READING

Note: Many of these resources are for the MCMI-III rather than the MCMI-IV. Newer resources will become available for the MCMI-IV in the near future.

Choca, J. P., & Van Denburg, E. (2004). *Interpretive guide to the Millon Clinical Multiaxial Inventory*, 3rd ed. Washington, DC: American Psychological Association.

Jankowski, D. (2002). *A beginner's guide to the MCMI-III*. Washington, DC: American Psychological Association.

Millon, T. (2011). *Disorders of personality: Introducing a DSM/ICD spectrum from normal to abnormal*. Hoboken, NJ: Wiley.

Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory—IV (MCMI-IV) manual*. Minneapolis, MN: Pearson Clinical Assessments.

Millon, T., Millon, C., Meagher, S., Grossman, S., & Ramnath, R. (2004). *Personality disorders in modern life* (2nd ed.). Hoboken, NJ: Wiley.

Millon, T., & Bloom, C. (2008). *The Millon inventories: A practitioner's guide to personalized clinical assessment* (2nd ed.). New York, NY: Guilford Press.

Strack, S. (2008). *Essentials of Millon inventories assessment* (3rd ed.). Hoboken, NJ: Wiley.

NEO PERSONALITY INVENTORY

The NEO Personality Inventory, in two versions (NEO PI-R and NEO-PI-3), is a self-administered, paper-and-pencil test composed of 240 statements rated on a 5-point Likert-type scale, from “Strongly Disagree” to “Strongly Agree.” Both the NEO PI-R (R standing for Revised) and the NEO-PI-3 (Third Edition) have two alternate forms, for self-rating (Form S) and for peer, spouse, or expert rating (Form R). These alternate forms include the same sets of statements, but in Form R they are altered from the first person to the third person. The test can be administered either to individuals or groups. The NEO PI-R is used to evaluate individuals between the ages of 17 and 89, while the NEO-PI-3 can be used with individuals between the ages of 12 and 99. The two tests are the same, except that the NEO-PI-3 has replaced 37 items with new statements that were developed to be more easily understood and have better psychometric properties. However, both versions of the test continue to be available for use. The NEO items request information concerning an individual’s typical behavior patterns; usual feelings and opinions; and attitudes relating to the self, others, and situations. The results are presented on a continuum of five overarching personality dimensions and 30 more specific facet scales. The five personality dimensions are Neuroticism, Extraversion, and Openness (which are the basis for the test name NEO), as well as Agreeableness and Conscientiousness (which were added later but did not alter the name of the test).

The philosophical orientation of the NEO is based on a five-factor model (or “Big 5”) of personality traits that have been found to be relevant throughout different cultures. The five factors—Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness—include both normal and abnormal personality traits (Markon, Krueger, & Watson, 2005), with each being somewhat bipolar. That is, in general, the more extreme an individual scores on *either* the low or high end of any one of the five scales, the more likely it is describing some unhealthy personality traits. The Big 5 have been found to be universal to humans (McCrae, Terracciano, Members of the Personality Profiles of Cultures Project, 2005), and, as such, it has been translated into more than 50 languages. The Big 5 have been found to have “functional validity,” including cross-cultural relevance, clear meaning to a wide range of people, and a high degree of power in predicting behavior. This is not to imply that untrained persons can competently interpret the NEO, but rather that the test’s roots and original constructs are based on conceptions of human behavior held by most people in most cultures. It is up to the skilled clinician to go beyond these common constructs and into a more subtle, broad, and integrated description of the person. The major focus and concern of the NEO involve descriptions of normal human traits that strive to be relevant, understandable, and accurate in terms of behavioral predictions.

The NEO was originally developed by Costa and McCrae (1985) and revised in 1992 into the now-used NEO PI-R. There is also a brief version of the test, the NEO-FFI, which provides scores on the Big 5 domains but not on the facet scales that make up these domains. Although reviews of the test have been mixed, most reviewers describe it in favorable terms, especially based on the strength of the five-factor model. For example, Benson (2014), in the nineteenth edition of *Mental Measurements Yearbook*, praised the “large body of theoretical literature and empirical studies to support score interpretation when attempting to understand and describe personality using the NEO Inventories” (p. 480). However, much of the evidence for the NEO-PI-3 must be extrapolated from earlier versions of the test, which is generally reasonable, given the lack of drastic change from the NEO PI-R to the NEO-PI-3.

HISTORY AND DEVELOPMENT

The NEO was developed as an inventory to assess enduring personality characteristics in a normal population. McCrae and Costa’s first version, developed in 1978, included only the three personality domains of Neuroticism, Extraversion, and Openness to Experience, giving the instrument the name “NEO” as an acronym. The measure was developed specifically with a hierarchical structure, incorporating overarching domains with component facets (Costa & McCrae, 1995). Upon further research, including encompassing many different models of personality, from Jung’s types to Gough’s folk concepts (Costa & McCrae, 2014), the researchers decided that five factors were necessary to comprehensively describe personality. Thus, in 1987, the NEO PI added the domains of Agreeableness and Conscientiousness. This version of the measure had measures of each of the five domains as well as six facet scales for each of the three original domains (McCrae & Costa, 1987). In 1992, the full NEO PI-R was published (Costa & McCrae, 1992), with the full five domains and six facet scales for each. In the meantime, the NEO FFI was published in 1989.

Like the MMPI, the NEO scores are given a standard score (*T* score) with a mean of 50 and a standard deviation of 10. The NEO PI-R scales were originally standardized on a normative sample of 2,273 adult males and females assessed between 1989 and 1991, from which 500 males and 500 females were selected, having been screened for the validity checks and random responding and then generally matching the U.S. census projections for 1995 in age and ethnic/racial distribution. The major criticism of this normative sample was the relatively higher education level than the general population. The standardization for 2005’s NEO-PI-3 was based on 242 male and 258 female adolescents, ranging from ages 14 to 20 years, as well as 279 male and 356 female adults, ranging in ages from 21 to 91 years. The adolescent sample was stratified by age and race, although it was predominantly White, mostly from Pennsylvania, and had overrepresentation of high academic achievers (McCrae, Costa, & Martin, 2005). The adult sample was similarly predominantly White, mostly from Pennsylvania, and was moderately overrepresented by high academic achievers and affluent adults (McCrae, Martin, et al. 2005).

The NEO can be administered individually or in a group, using pencil and paper or computer administration, and can even be mailed out to an individual to complete.

Scoring can be accomplished by hand, using scannable answer sheets, or with computer software. It is widely used in many settings, including research, clinical settings, and a variety of nonclinical settings, such as for vocational assistance or placement.

RELIABILITY AND VALIDITY

In general, the reliability and validity studies on the NEO compare favorably with those done on other personality inventories. Reliability has been assessed through internal consistency (coefficient alpha), short-term test-retest reliability, and longer-term stability studies. The NEO-PI-3 test manual (McCrae & Costa, 2010) reported internal consistency for the Big 5 personality domains as excellent, ranging from .89 to .93 for Form S and slightly higher for Form R. These are comparable to the alphas found for the Big 5 personality domains on the NEO PI-R, which ranged from .88 to .92 (McCrae, Martin, et al., 2005). Additionally, for adolescents assessed with the NEO PI-R, internal consistency for the Big 5 personality domains ranged from .87 to .95 (McCrae, Martin, et al., 2005). Short-term (1 week) test-retest reliabilities for the Big 5 personality domains on the NEO PI-R ranged from .91 to .93 (Kurtz & Parrish, 2001). Longer-term stability has been similarly found to be excellent, ranging from .78 and .85 for the Big 5 personality domains (Terracciano, Costa, & McCrae, 2006).

Reliability information for the 30 facet scales has been more mixed than the excellent reliability for the Big 5 personality domains. Although most have good internal consistency (a median of .76; McCrae & Costa, 2010), several fall below .70. The only truly problematic facet scale is O4: Openness to Actions, which has an internal consistency of .54 (McCrae & Costa, 2010). However, McCrae and Costa defended this facet scale as including extremely varied, nonoverlapping item content, even though the scale's items map onto the same trait. And, in fact, research on this facet scale has shown that it is generally comparable to other scales, and psychometrically sound, in stability, heritability, consensual validity, and cross-observer agreement (Jang, McCrae, Angleitner, Riemann, & Livesley, 1998; McCrae, Kurtz, Yamagata, & Terracciano, 2011; McCrae, Martin, et al., 2005; Terracciano et al., 2006). Short-term test-retest reliability was found to be acceptable for the facet scales (ranging from .70 to .91), with O4: Openness to Actions being .78 (McCrae & Costa, 2010). Stability was similarly adequate for the facet scales, ranging from .57 to .82, with a median of .70 (McCrae & Costa, 2010). Overall, reliability of both the Big 5 personality domains and the 30 facet scales is adequate to good.

Much research has focused on validating the five-factor model (Big 5) as a construct, outside of the NEO or other assessment tools. Markon, Krueger, and Watson (2005) conducted both a meta-analysis and an empirical study to determine if measures, and consequently theories, of both normal and abnormal personality functioning would converge under an umbrella of personality functioning. They found that among five different inventories, which covered multiple models of personality functioning, the most suitable overarching theory was a five-factor model, and the NEO PI-R was among the most successful inventories at identifying the specific factors.

Factor analytic studies have been employed to contribute to the validity evidence for the NEO. McCrae and Costa (2008) found that when factor analyzing the 240 items

of the NEO-PI-3, a five-factor solution was appropriate, and the calculated domain scores correlated very highly with the expected factors, ranging from .84 to .95. As expected, all correlations between the calculated domain scores and the factors other than the expected factors were below .34 (McCrae & Costa, 2010). Further, they found that the individual facet scales correlated most highly with the expected factors, further providing evidence that the facets do indeed cluster around their Big 5 domains. Further, research has revealed that the five factors can be reliably replicated for many different populations, including college students, adolescents, men, women, White and racial minority Americans, and across different cultures (Costa, McCrae, & Dye, 1991; McCrae, Terracciano, et al., 2005).

In addition to factor analysis, another important piece of evidence for validity is related to cross-observer agreement. When self-report and other reports significantly align, there is evidence that a measure is evaluating what it purports to evaluate. In general, cross-observer agreement of both the NEO PI-R and the NEO-PI-3 have been found to be adequate to good, ranging generally between .35 and .65 (generally, anything above .30 is considered adequate; McCrae & Costa, 2010). In a study of the Big 5 personality domains, McCrae (2008) found correlations between self-report and other report (comparing Form S and Form R) ranging from .56 and .67, again excellent for a personality measure. This provides evidence of the validity of the domain scores themselves, as well as evidence that bias on the part of self-report and other report may not affect the NEO too dramatically.

Finally, in order to establish validity, many studies have been conducted using the NEO to correlate with other criterion measures and outcomes. McCrae and Costa (2010) included information on how each facet scale correlates with five other criterion measures, which include scales from the Minnesota Multiphasic Personality Inventory (MMPI), the State-Trait Personality Inventory (STAI), the Interpersonal Style Inventory, and others. All correlations are statistically significant at the $p < .001$ level and converge on expected constructs. For example, facet scale N5: Impulsiveness has correlations of $-.46$ with the Revised California Personality Inventory Self-Control scale, $-.43$ with the Guilford-Zimmerman Temperament Survey Restraint scale, and $.43$ with the MMPI Borderline scale, all of which provide evidence for the construct of impulsiveness being appropriately measured on the NEO. McCrae and Costa (1992) also identified the highest correlating adjectives from Gough and Heilbrun's (1983) Adjective Check List with each of the 30 facet scales, and the adjectives were judged to discriminate the facet scales validly, so much so that raters could most often identify the exact nature of the facet scale simply from the list of correlated adjectives.

Other researchers have successfully correlated both the Big 5 domain scales and the 30 facet scales with scales of the MMPI (Hathaway & McKinley, 1983; Siegler, Zonderman, Barefoot, Williams, Costa, & McCrae, 1990), the Personality Assessment Inventory (PAI; Morey, 1991), and the Millon Clinical Multiaxial Inventory (MCMI; Lehne, 2002), again showing both that the NEO scales are likely measuring what they purport to measure and that the NEO measures both normal and abnormal personality functioning. Further, Hopwood, Flato, Ambwani, Garland, and Morey (2009) found that the NEO significantly predicted a range of social and behavioral outcomes, including dysfunction in relationships, work, and participation in pleasurable parts of life. Others have found significant correlations with a range of criteria, including

general psychological well-being (Steel, Schmidt, & Shultz, 2008), vocational interests (De Fruyt & Mervielde, 1997), attachment styles (Shaver & Brennan, 1992), and even defense mechanisms (Costa, Zonderman, & McCrae, 1991). An impressive amount of empirical literature has been amassed to link the NEO scales with appropriate criteria, including other measures and behavioral and clinical outcomes, building confidence that the measure is a valid measure of personality traits.

ASSETS AND LIMITATIONS

The NEO focuses on understanding personality traits in normal populations. Instead of focusing on pathology, it assesses areas on continua that represent a broad spectrum of styles, attitudes, values, thoughts, beliefs, and behaviors. However, even though its emphasis is on assessing normal variations, extreme scores can provide important information about the specifics of a person's expression of maladjustment, particularly with regard to personality pathology (Costa & McCrae, 1990; S. K. Reynolds & Clark, 2001; Wiggins & Pincus, 1989). In contrast to the MMPI and MCMI, which are limited to use with primarily pathologically oriented populations, the NEO is appropriate for normal persons. Thus, it addresses issues that interest a great many people.

In addition to its applicability to both pathological and nonpathological populations, the NEO has applicability in many varied settings. It has been successfully applied to clinical, educational, occupational, and even medical settings. The measure has been used for treatment planning (e.g., Sanderson & Clarkin, 2002), personnel selection (e.g., Black, 2000; Griffin, Hesketh, & Grayson, 2004), school counseling (e.g., Scepansky & Bjornsen, 2003), and in areas of behavioral medicine (Costa & McCrae, 2003), including predicting risk for certain diseases (Costa, Stone, McCrae, Dembroski, & Williams, 1987) and tracking changes in personality from chronic illnesses (M. E. Strauss & Pasupathi, 1994). Because it focuses not only on pathological aspects of personality functioning, even feedback from scales within normal limits can be useful in a variety of settings with a variety of individuals.

A major asset of the NEO is the fact that the language used even in the scoring and interpretation is relatively easily understood by a wide range of persons. This is not to say that anybody can accurately interpret the test, but it takes much less translation from scales to feedback and report than many other tests. Characteristics such as *modesty*, *warmth*, and *assertiveness* are generally straightforward and are, therefore, not easily misinterpreted by untrained professionals. In contrast, providing feedback to clients who have taken the MMPI requires the clinician to rephrase psychiatric terminology into more approachable, easily understood terminology. The language used in the questions, as well, is easily understood and generally nonthreatening, without inherent judgment in any given item. Additionally, it has been noted that feedback from the NEO, whether given verbally, in written form, or on the plotted charts, can be easily and effectively used for therapeutic feedback (Blonigen, Timko, Jacob, & Moos, 2015; J. A. Singer, 2005).

A further asset of the NEO relates to its nonthreatening focus on normative personality traits, rather than just pathology. Because of this nonthreatening nature, it lends itself not only to providing useful information, but also to building rapport

(T. Miller, 1991). Both the individual items and the results are easily discussed between the assessor and the individual being tested, most often with the ability not to focus on “bad news” or explaining specific pathology.

The major limitation of the NEO is the fact that there is no direct assessment of problems or problem behaviors (Ben-Porath & Waller, 1992). This is a limitation only in that the measure cannot be used as a stand-alone measure in clinical assessment. However, no measure should be used in the absence of other measures, and the NEO does not purport to assess specific psychopathology. As such, it is important to understand what the NEO can provide and what it cannot. In the absence of other measures, it will not provide specific diagnostic recommendations, although some scale configurations may be linked with certain types of pathology.

A further limitation of the NEO is its lack of solid validity scales. As a result, assessors do not have a formal means of determining if there have been attempts to fake good, fake bad, or respond randomly. In addition, the meaning of the individual items is relatively transparent and face valid, which makes it easier for respondents to respond misleadingly. Although much research has shown that the test is valid, the lack of scales or items to modify other scales and the transparency likely makes the test easier to manipulate. That is, individuals could skew test results relatively easily by manipulating their responses, if they so desire. Clinicians should be especially aware of this issue in situations in which motivated response alteration is likely, such as personnel selection or many forensic situations.

The NEO, then, is an extremely useful test in the assessment of personality characteristics of relatively normal persons, as well as in clinical assessments. It measures variables that interest a great number of people and uses relatively understandable concepts that lend themselves toward therapeutic feedback. Although purposeful fabrication of responses remains a potential issue, the measure has distinguished itself as an extremely useful tool in the evaluation of nonpathological populations, as well as a component in a larger clinical assessment.

USE WITH DIVERSE GROUPS

The NEO Personality Inventory has been translated into more than 50 different languages and dialects. Research on the five-factor model, predominantly using the NEO PI-R, has shown significant support for the emergence of the same five factors across different cultural groups (see McCrae, 2002). In general, the same five-factor structure that has emerged in American samples has emerged quite clearly in Belgian, Hungarian (De Fruyt, McCrae, Szirmák, & Nagy, 2004), Estonian (Kallasmaa, Allik, Realo, & McCrae), French (McCrae, Costa, Del Pilar, Rolland, & Parker, 1998), and Korean samples (Piedmont & Chae, 1997; Yoon, Schmidt, & Ilies, 2002). With Chinese and sub-Saharan African samples, although the internal consistency of the five factors was significantly weaker, the test-retest reliability was strong and a similar five-factor structure emerged (Piedmont, Bain, McCrae, & Costa, 2002; J. Yang et al., 1999). Age differences in personality structure were found to be similar across multiple cultures as well, including Germany, Italy, Portugal, Croatia, and South Korea (McCrae, Costa, et al., 1999). Also consistent across cultures is the finding that women tend to score

higher on Agreeableness than men and that people from European backgrounds (such as Americans, Australians, and Canadians) tend to score higher on Extraversion than people from Africa or Asia (Hofstede & McCrae, 2004).

Although the major factors were generally replicated in Filipino samples (Katigbak, Church, & Akamine, 1996; McCrae et al., 1998), closer scrutiny of items found some differential item functioning between Filipino and American college students (Huang, Church, & Katigbak, 1997). That is, about 40% of the items on the NEO had different psychometric properties between the two groups. However, it was concluded that the factors overall were still applicable within the Filipino culture. Cheung et al. (2001) found that the five-factor model was applicable in Chinese culture, but it was missing a qualitatively distinct, important aspect of Chinese personality, interpersonal relatedness.

In general, the preceding research supports the cross-cultural application of the five-factor model and more specifically the use of the multiple translations of the NEO Personality Inventory. However, there has been less focus on differential item functioning and differential trait functioning studies across cultures, as well as within-culture differences (e.g., potential differential functioning by socioeconomic status or gender identity). Much of the early work on the five-factor model, though, accounted for these within-culture differences.

INTERPRETATION PROCEDURE

Interpretation of the NEO Personality Inventory includes interpreting individual *T* scores of both the Big 5 domains and the facet scores individually, as well as thinking about them taken together. All the scales on the NEO are thought to approximate a normal, bell-shaped distribution, with fewer individuals having extremely low or extremely high scores on any scale. Each personality characteristic, including both the Big 5 and the 30 facet scores, represents a continuum of traits; that is, the continuous scale scores on each scale represent *degrees* of each trait, with those scores farther from the mean representing more distinctive features. A general description of the individual can be interpreted from the five factors, and more detailed information can be gleaned from the facet scales. While it is most common for the six facets that make up a factor to align consistently with the factor itself, special note should be taken when a facet score does not align with the overall factor. For example, if an individual scores high on Agreeableness (A), also scoring high on most of its facets but scores low on A3: Altruism, it can be an important distinguishing characteristic of that individual. He or she may overall believe in the good of others, maintain a humble and courteous manner, and respect others, but the person may not actively go out of his or her way to help others, especially when it is inconvenient to do so. The facets help specify and characterize the main factors.

In addition to individual domains being the focus of attention, pairs of domains can be particularly useful in characterizing different personality styles (Costa & Piedmont, 2005). The 10 styles, consisting of the different pair combinations of domains, include styles of interests, interactions, well-being, defense, anger control, impulse control, activity, attitudes, learning, and character. These styles further characterize, in useful terms,

individuals assessed using the NEO. Taken together, individual scales, facet-domain comparisons, and paired domain styles constitute the basis of interpreting the NEO.

Individual Scales

There are no specific validity scales within the NEO-PI-3 or NEO PI-R. Research has shown that response distortion is not a major problem when assessing generally normally functioning adults or in research settings with a degree of anonymity (Piedmont, McCrae, Riemann, & Angleitner, 2000; J. Yang, Bagby, & Ryder, 2000). However, response distortion may be a bigger issue in certain circumstances, such as personnel selection (Young & Schinka, 2001), when motivated distortion is likely. Rather than including a validity scale with items, there are a few validity checks for the measure. First, as part of the measure, there are three self-report validity questions. Question A asks if the individual has responded honestly and accurately. Question B asks if the individual has responded to all items, and question C asks if the individual has marked his or her responses in the appropriate spaces. Typically, responding *disagree* or *strongly disagree* to question A would preclude the scoring and interpretation of the test. Responding *No* to questions B and C can be addressed and corrected in the moment; if these issues are not addressed, again the test may be rendered invalid. Specifically, if more than 40 items are missing from a protocol, it should not be scored and interpreted.

In addition to these three validity questions, there are several other steps to increase confidence in a valid profile from the NEO. First, to address response bias (either agreeing excessively or excessively not agreeing), the total number of *agree* and *strongly agree* responses should be counted. Costa, McCrae, et al. (1991) found that 99% of respondents agreed with more than 50 items and fewer than 150 items. Thus, if the total number of *agree* and *strongly agree* responses is less than 51 or more than 149, the results should be interpreted with caution.

Random responding is difficult to identify in the NEO; however, some guidelines are provided to look for likely carelessness or resistance. McCrae and Costa (2010) recommended considering invalid any protocol that includes certain numbers of consecutive strings of the same response. Specifically, endorsing SD (*strongly disagree*) to more than 6 items in a row, D (*disagree*) to more than 9 items in a row, N (*neutral*) to more than 10 items in a row, A (*agree*) to more than 14 items in a row, or SA (*strongly agree*) to more than 9 items in a row should be considered to invalidate the test. Additionally, Schinka, Kinder, and Kremer (1997) developed an inconsistency scale (*INC*) consisting of 10 pairs of items that are overwhelmingly consistent in their relationship. The item pairs include: 11 and 71, 39 and 159, 53 and 113, 59 and 199, 72 and 132, 85 and 215, 102 and 162, 110 and 170, 188 and 218, and 191 and 221. An *INC* score is calculated by taking the difference in each paired score and summing them. Protocols with a raw *INC* score of 10 or higher should be examined to determine the likelihood of random responding. The *INC* scale has gained some research support for discriminating between random responders and their counterparts (Scandell, 2000; Young & Schinka, 2001).

In addition to random responding, Schinka and colleagues (1997) created scales to represent negative presentation management (*NPM*) and positive presentation

management (*PPM*), to evaluate responders who are deliberately trying to present themselves in either a negative or positive light. The *NPM* is calculated by summing the following scores: for items 15, 48, 57, 62, 73, 104, 129, and 135, a response of SD gets 4 points, D gets 3, N gets 2, A gets 1, and SA gets 0 points. For items 31 and 161, a response of SD gets 0 points, D gets 1, N gets 2, A gets 3, and SA gets 4 points. Generally, if the sum of the scores assigned to these 10 items is 15 or higher in the general population (or 11 or higher in personnel selection, where negative presentation is rare), the protocol should be evaluated more closely for possible deliberate negative impression management. For *PPM*, the scale is calculated by summing the following scores: for items 30, 42, 113, 146, 162, and 196, a response of SD gets 4 points, D gets 3, N gets 2, A gets 1, and SA gets 0 points. For items 37, 93, 139, and 153, a response of SD gets 0 points, D gets 1, N gets 2, A gets 3, and SA gets 4 points. Generally, if the sum of the scores assigned to these 10 items is 25 or higher in the general population (or 34 or higher in personnel selection, where some positive impression management is expected), the protocol should be evaluated more closely for possible significantly deliberate positive presentation. The general cutoff scores are based on the infrequency of individuals who tend to achieve those scores (Ballenger, Caldwell-Andrews, & Baer 2001; Schinka et al., 1997), and again these scales have found positive support in the literature (Caldwell-Andrews, Baer, & Berry, 2000; Griffin et al., 2004; C. A. Pauls & Crost, 2005; Scandell, 2000; Young & Schinka, 2001). It should be noted that although these validity scales have some support, the support for them in actual clinical settings is not as good (Morey, Quigley, et al., 2002; Piedmont et al., 2000; J. Yang et al., 2000), so clinicians should use them to scrutinize but perhaps not throw out protocols when they are deemed questionably valid.

Neuroticism (N)

The Neuroticism (N) scale measures the tendency toward emotional instability, turmoil, and general distress. This personality trait is thought to underlie a host of emotional difficulties, including anxiety, depression, anger, low self-esteem, impulsivity, and general malaise, among others. Individuals who score high on N ($T > 55$) tend to be anxious, emotionally labile, quick to anger, sad, and in general have disruptive emotions. They are likely to have low self-worth and question their own effectiveness. They may act impulsively without fully understanding why they do so, and they can easily become embarrassed and self-conscious, feeling shame and guilt easily. Individuals who score low on N ($T < 45$) tend to be more emotionally stable and secure, even-tempered, and less prone to sadness, loneliness, and embarrassment. They tend to be less impulsive. However, individuals who score extremely low on N may be less productive, without the necessary baseline level of anxiety to provide a sense of urgency.

The Anxiety facet scale (N1) measures proneness to becoming tense, jittery, nervous, and fearful in general. Although the scale does not measure specific instances of fears or phobias, those who score high on N1 are more likely to experience these specific anxieties. Individuals who score high on N1 ($T > 55$) tend to be skittish, easily frightened, and generally apprehensive about their situations and the future. They may ruminate about things that have gone wrong or may go wrong in the future. Individuals who score low on N1 ($T < 45$) have fewer fears and apprehensions than most other people; they likely do not dwell on things that may go wrong in their lives. They tend

to be even-keeled and calm under pressure, although at extremely low levels of N1 they may not have enough anxiety to prod them to be productive.

The Angry Hostility facet scale (N2) measures the tendency toward anger, bitterness, and resentment. It is not a measure of aggression (behavior), which will likely depend on other personality characteristics, like Agreeableness. Individuals who score high on N2 ($T > 55$) tend to be quick to anger, often described as having a short fuse, and are easily disgusted by others and become bitter and resentful. These individuals often become frustrated and angry at even minor annoyances, and they may adopt a hostile attitude toward others around them. Individuals who score low on N2 ($T < 45$) tend to be much more even-tempered and resistant to annoyances and frustrations. Individuals who score extremely low on this scale may appear less passionate about circumstances in their lives, and these individuals are likely to score high on Agreeableness.

The Depression facet scale (N3) measures the likelihood of individuals experiencing the range of depressive affects, including sadness, loneliness, hopelessness, helplessness, worthlessness, guilt, shame, and others. Individuals who score high on N3 ($T > 55$) tend to be sad, gloomy, pessimistic, and self-deprecating. They tend to attribute negative events to personal, internal causes and to believe in the stability of negative circumstances and the fleeting nature of positive ones. Individuals who score low on N3 ($T < 45$) are less likely to feel worthless, helpless, hopeless, or lonely. They likely attribute failures to external causes rather than to personal ones, and they tend not to feel a deep sense of shame or guilt. These individuals may not necessarily appear cheerful and carefree, however, as this presentation depends also on Extraversion.

The Self-Consciousness facet scale (N4) is a uniquely social component of the Neuroticism domain. N4 measures discomfort with social awkwardness, especially focused on feelings of shame and embarrassment. Individuals who score high on N4 ($T > 55$) tend to become anxious in situations in which they feel they are being judged or evaluated by others. They are sensitive to being teased, become easily embarrassed or ashamed in social situations (even when it is people they know, rather than themselves, who make mistakes), and constantly fear making social mistakes. Individuals who score low on N4 ($T < 45$) may not make fewer social mistakes than their peers, but they are less likely to be worried or bothered by them.

The Impulsiveness facet scale (N5) is not a measure of risk taking or hasty decision making. Rather, on the NEO, N5 measures the degree to which individuals have difficulty resisting their urges. Individuals who score high on N5 ($T > 55$) tend to give in to their urges without a high degree of self-restraint. They may overindulge in food or other cravings, and they tend to have difficulty keeping control over their emotions. They often exhibit behaviors that they later regret. Individuals who score low on N5 ($T < 45$) do not find their cravings difficult to resist, and as such they tend not to overeat or overindulge in ways that make them sick or they later regret. They tend to have an easier time maintaining control over their emotions.

The Vulnerability facet scale (N6) measures the degree to which individuals feel capable or not of coping with stress. Individuals who score high on N6 ($T > 55$) feel as if they do not have the resources to cope with stress and are worried that they will fall apart without depending on others and feeling hopeless and worried when difficult situations present themselves. They feel helpless and emotionally insecure and often

have difficulty being decisive. Individuals who score low on N6 ($T < 45$) feel more able to handle even extreme stressors and difficult situations. They feel they can make good decisions even when under pressure, and they are generally emotionally stable and calm under stress.

Extraversion (E)

The Extraversion (E) scale measures the degree to which individuals are not only outgoing and sociable, but also assertive, upbeat, warm, and friendly. This personality trait includes a degree of positive emotionality and a willing and enthusiastic engagement with the social world. Individuals who score high on E ($T > 55$) tend to be talkative, warm, and friendly. They tend to be leaders, because of natural assertiveness, and at higher levels even can be quite socially dominant and even aggressive. They are energetic and generally present with high spirits, and they strongly prefer situations in which there are many people around them. Individuals who score low on E ($T < 45$) tend to prefer being on their own. They can be quite reserved in social situations, but this is usually not due to social anxiety. They tend to think and work in a more even-paced manner, though not necessarily more slowly. They do not appear as effervescent as their high-E counterparts, though they are not necessarily pessimistic or unhappy.

The Warmth facet scale (E1) measures comfort with interpersonal intimacy and closeness. Individuals who score high on E1 ($T > 55$) are genuinely interested in interacting with others and developing close, intimate relationships. They will appear to others as compassionate, affectionate, and outgoing, easily building emotional bonds and closeness with others in their lives. Individuals who score low on E1 ($T < 45$) tend to adopt a more formal, distant stance toward others in their lives. They are not necessarily unfriendly, but they are more reserved when relating to others, which others can interpret as being cold or detached.

The Gregariousness facet scale (E2) measures the preference for having other people around. Individuals who score high on E2 ($T > 55$) strongly prefer to have others around and greatly enjoy other people's company. Their longing for other people grows when they are alone for an extended period of time. Individuals who score low on E2 ($T < 45$) tend to be loners, preferring solitary activities to those with lots of other people around. When given a choice, they will likely choose activities they can do alone rather than group activities, and they may actively avoid situations with large groups of people.

The Assertiveness facet scale (E3) measures the tendency to make oneself heard and known in social situations. Individuals who score high on E3 ($T > 55$) stand up for themselves easily and often take leadership positions when available. They can be dominant and forceful in some situations, forcing their opinions on others. Individuals who score low on E3 ($T < 45$) prefer to be followers rather than leaders, often letting other people assert themselves rather than talking in social situations. They tend to take a passive stance in group decision making, and, even when they may be the most qualified, they find it difficult to step up and take on a leadership position.

The Activity facet scale (E4) measures the amount of energy and gusto with which individuals live their lives. Individuals who score high on E4 ($T > 55$) live fast-paced lives, engaging vigorously in activities and bringing a sense of energy and verve to

situations. They may seem to always be in a hurry or extremely busy. Individuals who score low on E4 ($T < 45$) are more relaxed in their general demeanor, more leisurely in the way they approach situations and activities in their lives. They may appear to have less of a sense of urgency in general, less pressured to keep moving and working, though they are typically not lazy or lethargic.

The Excitement-Seeking facet scale (E5) measures the need for and enjoyment in high-stimulation activities. Individuals who score high on E5 ($T > 55$) crave excitement and stimulation, often participating in activities that are thrilling and high energy. Individuals who score low on E5 ($T < 45$) do not have a need for high levels of excitement in activities and are content with the more mellow aspects of life. They tend to avoid situations that are overly stimulating.

The Positive Emotions facet scale (E6) measures the tendency to experience positive emotions like happiness, joy, bliss, and love. Individuals who score high on E6 ($T > 55$) laugh easily, are jovial and optimistic by nature, and often feel excited. Individuals who score low on E6 ($T < 45$) may not necessarily be unhappy, but their experience of positive emotions is less exuberant and enthusiastic. They do not tend to experience or express the general lighthearted excitement that those with high E6 do.

Openness (O)

The Openness to Experience (O) scale encompasses many different personality traits that have been researched in the literature, including imagination, curiosity, attunement toward personal emotions, and preference for abstract thinking. The curiosity that is measured by this scale relates to curiosity about the self, others, and the world in general. Individuals who score high on O ($T > 55$) are intellectually and creatively curious, open to new ideas and to values and theories that may contradict or challenge their own. They are aware of and can tolerate and even enjoy feeling their emotions more than others. They enjoy playing around with ideas, theories, and problems and may appear unconventional or quirky in their behavior and worldview. Individuals who score low on O ($T < 45$) are more conservative and conventional in the way they approach the world. They are not as attentive to their emotions (positive or negative), and they tend to be very realistic and level-headed about solving problems. They prefer a world that is predictable and familiar, rather than new and exciting, and they are likely to be more politically conservative. It should be noted that those who score low on O are not necessarily actively intolerant of others who are different or who think differently from them, though they often prefer those who are similar.

The Fantasy facet scale (O1) measures how active one's imagination is and how much fantasy is used not as an escape from reality, but as a way to create, solve problems, and even interact with the world. Individuals who score high on O1 ($T > 55$) have extremely active imaginations and a rich, vivid fantasy life. They use all facets of their fantasy world to create imaginative solutions to problems. Individuals who score low on O1 ($T < 45$) prefer their thoughts to be linear and grounded in reality, not finding daydreaming or fantasy exploration useful. They tend to keep a very good focus on tasks at hand, not allowing their minds to wander unnecessarily.

The Aesthetics facet scale (O2) measures the interest in and absorption by art, beauty, and even the inherent beauty in nature. Individuals who score high on O2 ($T > 55$) have a deep appreciation in the arts and are often moved deeply emotionally

by music, poetry, dance, and other art forms. They appreciate the patterns and inherent beauty in both nature and art and can become engrossed in the emotions they feel when experiencing art. They may not be artistically talented or have what would be considered to be discerning taste, but they become truly absorbed in experiencing artistic works. Individuals who score low on O2 ($T < 45$) are less sensitive to, interested in, and feel little excitement from experiencing art. They likely become bored by artistic expression and tend not to value the beauty or emotional depth of works of art.

The Feelings facet scale (O3) measures openness to one's inner emotional life, including both allowing oneself to feel deep emotions and valuing them as integral to the entire, full human experience. Individuals who score high on O3 ($T > 55$) feel both positive and negative emotions more deeply than others, are able to differentiate them easily, and appreciate the depth of their impact on them. Because they feel emotions deeply and value their importance in life, they tend to empathize with others more easily, and they can easily connect with others on an emotional level. Individuals who score low on O3 ($T < 45$) have much less variation and depth of experience of their emotions than others. They tend to have flatter emotional experiences, and they often do not even recognize emotions in the moment. They may not pay attention to their emotional lives, not feeling that emotions are important, or they may not even have the ability to recognize, understand, or even feel emotions.

The Actions facet scale (O4) measures the behavioral aspects of openness, relating to trying novel and unknown activities, foods, and places. Individuals who score high on O4 ($T > 55$) prefer variety and become excited by trying new things, traveling, and eating new foods. They get bored with routine easily, and they may appear to make impulsive decisions, simply to change their routine. Individuals who score low on O4 ($T < 45$) much prefer the familiar and routine and even find change difficult. They stick to known and understood methods for achieving goals and likely find unfamiliar surroundings intimidating.

The Ideas facet scale (O5) measures a cognitive aspect of openness, related to intellectual curiosity. O5 incorporates aspects of genuine curiosity to understand concepts and ideas, as well as the openness to considering new, competing, unconventional, and even paradigm-shifting ideas. Individuals who score high on O5 ($T > 55$) are curious about the world around them and often other people. They enjoy playing with abstract ideas, using analogies to solve problems, and engaging in philosophical discussions about grand ideas and questions. Individuals who score low on O5 ($T < 45$) have little interest in engaging in philosophical debates about the grand questions in life, having limited curiosity about the world around them. They may be intellectually engaged in certain topics, but these tend to be narrow and highly focused. These individuals would rather spend their mental energy on concrete, specific thoughts rather than abstract, expansive ones.

The Values facet scale (O6) measures a willingness to reexamine one's values in the face of people with different worldviews, different cultures, and a changing world in general. Individuals who score high on O6 ($T > 55$) are open to accepting the differing values and lifestyles of others around them. They tend to be cultural relativists, believing that what is appropriate and acceptable in one culture may not necessarily hold true in another. Similarly, they tend not to believe in absolute right versus wrong,

feeling that situations, circumstances, and context are important factors when determining the merit of others' behaviors. Individuals who score low on O6 ($T < 45$) tend to be much more traditional, valuing and accepting authority and tradition. They are generally more conservative in their beliefs, typically believing that morality is not as much of a gray area as society has let it be. They value their own clear-cut definitions of right and wrong and feel that being true to their own principles is more important than being accepting of others.

Agreeableness (A)

The Agreeableness (A) scale measures both attitudes about the trustworthiness and general goodness of others and behaviors related to respecting, empathizing with, and deferring to others. Individuals who score high on A ($T > 55$) tend to be sympathetic toward others, willing and eager to help, and generally cooperative, believing that others are generally decent and honest. They try to be considerate to others, taking others' feelings and concerns into account whenever possible. They genuinely believe that others are just as caring and well intentioned as they are, expecting cooperation, help, and kindness in return for their own. These individuals can be seen as passive, compliant, and even weak and dependent. Individuals who score low on A ($T < 45$) are more skeptical of others' intentions, expecting competition and challenge from people around them. They tend to push to get their own way, though this can take different forms. Some will be outright antagonistic and aggressive to achieve their goals, while others may use manipulation and trickery to get others to do what they want. These individuals can be sarcastic and stubborn, digging in their heels when they want something done a certain way. They can be effective leaders when strong leadership is necessary.

The Trust facet scale (A1) measures faith in the goodness of the human spirit. Individuals who score high on A1 ($T > 55$) believe that others are well-intentioned and trustworthy people, honest even when not in their own best interest. They generally assume the best about people. Individuals who score low on A1 ($T < 45$) are more cynical about the goodness of humanity, believing that others may be dishonest and take advantage of them if given the chance. Rather than accepting and appreciating when someone does something nice for them, they can become suspicious about the intentions of the other person.

The Straightforwardness facet scale (A2) measures directness, unflinching honesty, and especially genuineness. Individuals who score high on A2 ($T > 55$) are sincere and genuine, regardless of the situation. They avoid hypocrisy in their own behaviors, and they tend not to be crafty in their interactions with others. Individuals who score low on A2 ($T < 45$) are more likely to use flattery and deception to get their needs met, even manipulating others when they feel it is in their own best interest. These individuals are not necessarily dishonest, but they are shrewder in their interpersonal dealings than their peers.

The Altruism facet scale (A3) measures genuine concern for the well-being of others. Individuals who score high on A3 ($T > 55$) tend to be courteous, generous, and considerate toward those around them. They often go out of their way to be helpful to others. Individuals who score low on A3 ($T < 45$) are more concerned with their own

well-being than that of others. They are reluctant even to become involved in other people's problems, and they tend to be more self-focused and even cold toward others.

The Compliance facet scale (A4) is specifically related to the way individuals react to conflict with others. A4 measures the degree to which individuals inhibit their own anger and defer to others. Individuals who score high on A4 ($T > 55$) tend to ignore their own angry feelings so as not to become aggressive, deferring to others in conflict. They would rather cooperate with others and preserve relationships than compete, even when they feel they are correct. They are likely to forgive and forget, even when others directly insult them. Individuals who score low on A4 ($T < 45$) often prefer to be right than to preserve a relationship, having no problem expressing their anger and even being aggressive. They can be stubborn and sarcastic, and they do not back down from a fight. Their interpersonal relationships, especially with family and coworkers, are often characterized by arguments and fights.

The Modesty facet scale (A5) is not a measure of self-esteem but measures the outward-facing trait of humility, including beliefs about being better than or equal to others and whether they would be likely to boast about their achievements. Individuals who score high on A5 ($T > 55$) generally believe they are no better than anybody else, regardless of their personal circumstances, and as such work hard not to boast about accomplishments. They are humble and prefer praising others to being praised themselves. Individuals who score low on A5 ($T < 45$) feel superior to others and need others to acknowledge that they are in fact superior. They tend to brag about their accomplishments to the extent that they are often considered arrogant and narcissistic.

The Tender-Mindedness facet scale (A6) measures sympathy and concern for others. Individuals who score high on A6 ($T > 55$) believe that social policy should be more concerned with the welfare of people rather than other, more impersonal factors like economics. They believe that all people deserve respect and have sympathy for those who are less fortunate than they are. Individuals who score low on A6 ($T < 45$) are not as concerned with individual people, tending to be more realistic, rational, and logical about decisions and beliefs than moved by pity. Fairness and the greater good are generally more important than mercy for those who are less fortunate.

Conscientiousness (C)

The Conscientiousness (C) scale measures an array of traits related to both an orientation toward accomplishing things and the behavioral correlates of doing so successfully. Included in these traits are determination and purposefulness as well as organization, the ability to prepare and execute plans, and skills needed to ultimately achieve in academic, occupational, or other pursuits. Individuals who score high on C ($T > 55$) are both motivated to achieve their goals and have the planning and organizing tools to do so. They are clear and considered in their planning, making lists and goals logically with appropriate steps to complete them. They think actions through thoroughly before enacting them, taking into consideration the potential consequences for alternative choices. Although they have a drive and the tools to accomplish their goals, they can also be compulsive about their work and perfectionistic about details. They are extremely reliable, but they can also be moralistic and even judgmental of both themselves and others. Individuals who score low on C ($T < 45$)

are not methodical in the way they approach problems and are often unprepared and disorganized. They often do not think through potential consequences before acting, seeming impulsive, and they often fall short of success on individual tasks, because of poor planning, poor time management, or a general lack of meticulousness in the process. They are most often not overly driven to succeed, and thus they approach tasks in a much more laidback, easygoing manner.

The Competence facet scale (C1) measures the feeling that one is generally effective and capable to succeed in tasks and life in general. Individuals who score high on C1 ($T > 55$) feel that they are prepared to tackle situations, generally well informed, high in common sense and sensibility, and consequently successful in life. These individuals pride themselves on their sound judgment and effectiveness. Individuals who score low on C1 ($T < 45$) are less confident in their own abilities, feeling that they are often ill-prepared to face the demands of situations. As such, they feel they are inefficient and ineffective at accomplishing tasks.

The Order facet scale (C2) measures a preference for neatness, tidiness, and orderliness. Individuals who score high on C2 ($T > 55$) prefer things to be in just the right place, and are extremely well organized, neat, and clean, so that they know exactly where things are and in what condition. Individuals who score low on C2 ($T < 45$) care less about order, often misplacing things and needing to spend time looking for them. They tend to plan ahead less, preferring to keep their options open. They also tend to be less perfectionistic and exacting in their standards.

The Dutifulness facet scale (C3) measures the degree to which individuals do things that they feel they *should* do. These things relate both to activities that are generally expected of them by society, such as following through on plans, and to situations governed by their morals and ethics. Individuals who score high on C3 ($T > 55$) work hard to follow through on all tasks that are expected of them, doing as good a job as possible so as not to either need to do them again or to be seen as not pulling their own weight. They adhere strongly to their own moral and ethical principles and can thus be more judgmental of themselves and others than most. Individuals who score low on C3 ($T < 45$) are unreliable and cannot be consistently depended on. Additionally, they tend to have a much more casual attitude toward ethical and moral judgments.

The Achievement Striving facet scale (C4) relates to an attitude of aspiration and striving to succeed in their goals. Individuals who score high on C4 ($T > 55$) work hard to achieve all their goals in a very purposeful and diligent manner. They may be workaholics, sacrificing other parts of their lives to ensure that they succeed in achieving their goals. Individuals who score low on C4 ($T < 45$) lack ambition and have no drive or need to succeed. They are generally content with not achieving highly, and they may even be lazy or aimless.

The Self-Discipline facet scale (C5) relates to the ability to actually follow through on a plan once it is set. Individuals who score high on C5 ($T > 55$) set goals clearly and, regardless of distraction or boredom, set up processes to ensure that they complete the goals. They are motivated to complete even meaningless tasks, as long as doing so is expected of them (by themselves or others). Individuals who score low on C5 ($T < 45$) tend to procrastinate and get easily distracted from completing tasks, often starting on new projects in the middle of unfinished ones. They simply do not possess the motivation to follow through on tasks to completion.

The Deliberation facet scale (C6) measures the degree to which individuals think and plan out carefully before acting. Individuals who score high on C6 ($T > 55$) plan ahead carefully and think through the potential consequences of behaviors before enacting them. They are cautious in their actions, being careful to behave in specific and deliberate ways. Individuals who score low on C6 ($T < 45$) are more spontaneous and hasty in their decisions, often behaving in ways that do not consider the consequences of their actions. They tend to ask for forgiveness more than asking for permission.

Personality Styles

Pairs of domain scores are used for various purposes and are labeled as “styles.” These styles encompass various domains of functioning, from interests to impulse control to interpersonal interactions. They are represented on the NEO on 10 style graphs, which plot the score on one domain against the score of the other area of interest. For example, Style of Interests is represented as a graph that plots Extraversion (E) on the vertical (y) axis and Openness (O) on the horizontal (x) axis. Whichever quadrant the individual’s scores land in provides a style for that person, and the farther from the intersection point of the axes, the more ingrained the style is. Figure 10.1 presents an example of a Style of Interests graph.

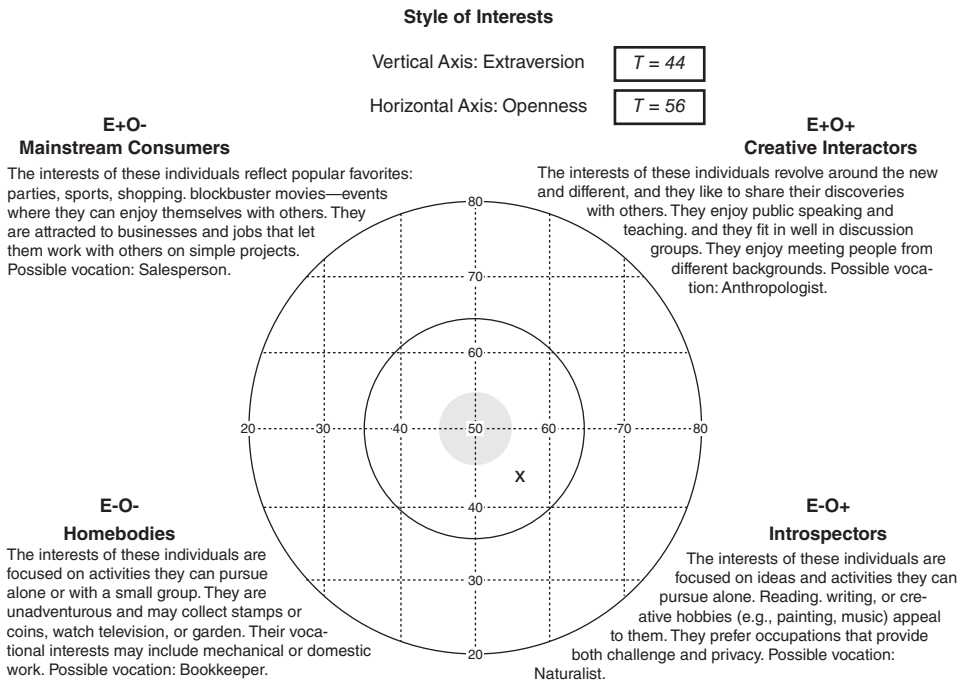


Figure 10.1 Example of Style of Interests graph

Source: Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, FL 33549, from the *NEO Inventories Professional Manual* by Robert R. McCrae, PhD, and Paul T. Costa Jr., PhD. Copyright 2010 by Psychological Assessment Resources, Inc. (PAR). Further reproduction is prohibited without permission of PAR.

Style of Interests

The Style of Interests category relates to preferences in the activities the individual engages in. It is composed of Extraversion and Openness, as these two continua help categorize general preferences in activities, whether solitary or group related, or familiar or novel.

Low E–Low O, the Homebody, represents individuals who prefer participating in activities that are both familiar and largely solitary. These individuals are likely to be found doing things like watching television or reading a great deal, as their desire for novel and invigorating activities, especially within groups, is minimal. These loners have little need for socialization, especially when that socialization occurs within novel contexts.

High E–Low O, the Mainstream Consumer, represents individuals who are outgoing and prefer to be around people but are largely interested in conventional and mainstream activities. Movies, shopping, and sporting events may be activities of choice for these individuals, as they feed their need for interaction without the pressure of novel and unfamiliar situations.

High E–High O, the Creative Interactor, represents individuals who enjoy exploring new and creative territory in life and especially love sharing those experiences with others. Traveling with groups is extremely appealing to these individuals, as is vigorous, friendly debating and discussion within large groups of people from different backgrounds and worldviews. They tend to be good teachers and public speakers, even in situations in which the audience is unknown.

Low E–High O, the Introspector, represents individuals whose interests lie in the kinds of thoughtful exploration they can do alone. Reading and artistic endeavors appeal to these individuals, as they can broaden their horizons in an independent way. They are likely drawn to activities in which they are challenged to be creative thinkers.

Style of Interactions

The Style of Interactions category relates to how individuals tend to treat social interactions with others and groups. It is composed of Extraversion and Agreeableness, as these two continua are both inherently social.

Low E–Low A, the Competitor, represents individuals who treat social interactions generally as competition, guarding themselves and wary of others as potential enemies or people out to take advantage of them. They tend to value their privacy above all else, keeping their motives and intentions private and hidden from others.

High E–Low A, the Leader, represents individuals who are decisive and unafraid to make their opinions and values known to others, as well as to give orders. They tend to make poor followers, but they can rally a team to work well together. At particularly high levels, these individuals are likely to be narcissistic and bossy, showing off for others but not taking others' perspectives into account very much, if at all.

High E–High A, the Welcomer, represents individuals who are both gregarious and warm toward others. They genuinely enjoy the company of others and are welcoming and empathic toward them. They are as prone to listen and sympathize as they are to give advice, and in general these individuals are warm, friendly, and likeable.

Low E–High A, the Unassuming, represents individuals who tend to be modest and relatively quiet in relationships. They prefer being alone, but they will easily go with the

will of others in a group. They are sympathetic and trusting, often not asserting themselves when their needs are not being met. These individuals are easily taken advantage of, and they tend to be very good team players, not allowing their personal interests or ideas interfere with the larger group's goals. However, when they have genuinely valuable input to add to projects, it is likely that they will not assert themselves enough to present it, and if they do, they are likely not to be heard and acknowledged as seriously.

Style of Well-Being

The Style of Well-Being category relates to the internal emotional world of the individual. It is composed of Extraversion and Neuroticism, as these two continua relate to the experience of generally positive and generally negative affectivity, respectively.

Low N–Low E, the Low-Keyed, represents individuals who are relatively unaffected emotionally by most circumstances and contexts. They feel neither high highs nor low lows, and in interpersonal relationships they are likely to be seen as somewhat stoic or cold. They do not tend to experience swings in mood, and their decisions are not likely to be made impulsively or to be heavily influenced by emotional factors.

High N–Low E, the Gloomy Pessimist, represents individuals who are characteristically sad, mopey, and blue most of the time. They have difficulty finding joy in life, and they are highly likely to become depressed. Generally pessimistic and often worried about life, these individuals can be very difficult to cheer up.

High N–High E, the Overly Emotional, represents individuals who feel the broad spectrum of emotions quite fully. At an extreme, they may be emotionally labile, shifting rapidly between different emotions, and may look histrionic, which can cause drama within relationships. At less of an extreme, they are highly attuned to their own emotional experience and are likely to have deep, empathic relationships with others.

Low N–High E, the Upbeat Optimist, represents individuals who are generally positive and optimistic, looking forward to the future and what it holds. They are cheerful, friendly people, and they recover from disappointment and anger quite quickly and easily. At an extreme, they may deny the negative or sadder aspects of life, experiencing only the joy and happiness of looking forward.

Style of Defense

The Style of Defense category relates to the general psychological strategies used to cope with difficulty in life, stress in general, and interpersonal conflict. It is composed of Openness and Neuroticism, focused on how individuals are likely to respond to high-stress situations.

Low N–Low O, the Hyposensitive defender, represents individuals who tend to focus on concrete action rather than the abstract notion of distress when difficulties arise. They are not easily moved to worry or concern, but they rather focus their energy on solving problems. They may feel that what happens is out of their control, so why would they bother to worry about it?

High N–Low O, the Maladaptive defender, represents individuals who tend to use primitive defense mechanisms to cope with difficulties in life. They may repress or suppress their feelings, deny that a situation is even occurring, and they typically have very little insight into their own role in problems. They will have negative emotional

reactions, such as depression or anxiety, without necessarily knowing or understanding where they may have originated.

High N–High O, the Hypersensitive defender, represents individuals who are prone to anticipating all of the possible things that can go wrong in any given situation. These individuals likely have vivid imaginations, and the way they put these imaginations to use is by creating sometimes difficult or even bizarre situations that can have negative consequences. These creative thoughts can plague these individuals, depending on how extreme on the scales they are, leading to rumination and even fantasized misfortune.

Low N–High O, the Adaptive defender, represents individuals who react to stressful situations by becoming somehow inspired, not allowing misfortune to overwhelm them entirely, but planning creative solutions to problems as they go. These individuals are indeed aware of the threats of misfortune in life, but they often use humor or artistic outlet to cope with the stress. They will adapt well to many different situations, even when under high pressure, as they do not allow negative emotions to interfere with creative problem solving.

Style of Anger Control

The Style of Anger Control category encompasses both the experience and expression of anger. It is composed of Neuroticism and Agreeableness, which in combination describe distinct strategies to feeling and expressing anger.

Low N–Low A, the Cold-Blooded, represents individuals who are calculated and strategic in their reaction to anger. They rarely appear overtly angry, and they often do not respond to situations that anger them immediately. Alternatively, they find ways to express their anger, often by means of revenge, in both a time and manner that suits them better than the immediate moment of offense. Their revenge may be manipulative and calculated and at extreme levels may even become criminal.

High N–Low A, the Temperamental, represents individuals who are hot-tempered and generally express their anger quickly and easily, even over a relatively minor offense. These individuals are not only quick to anger, but they hold on to their anger for long periods of time. They often do not do a good job of judging the potential consequences of expressing their anger, and they care more about voicing their anger than reflecting on the effect it may have on others around them. At higher levels, these individuals may become aggressive, both verbally and physically.

High N–High A, the Timid, represents individuals who are deeply ambivalent about their anger. They are easily hurt and offended, feeling slighted by even innocuous statements or jokes, but they tend not to want to express their anger for fear of hurting others. They may ruminate on their anger, and they often second-guess themselves for not asserting themselves in situations that anger them, blaming themselves for being too tentative. Their anger may get displaced onto other aspects of their lives or themselves.

Low N–High A, the Easy-Going, represents individuals who are not easy to offend or anger. They take insults or jokes in stride, feeling that very little will be gained by getting angry or expressing anger. If truly insulted, they will assert themselves, but they tend to do this in a calm, respectful way, rather than becoming overly emotional or aggressive. Favored tactics in difficult, angering situations are forgiving and forgetting or finding rational ways to broker peace and compromise.

Style of Impulse Control

The Style of Impulse Control category relates to the ability to maintain control over urges, desires, and behavior in general. It is composed of Neuroticism and Conscientiousness, to account for both the general tendency to become consumed with urges and the general style of behavioral control an individual tends to employ.

Low N–Low C, the Relaxed, represents individuals who are comfortable exerting very little control over their own behaviors, as they are generally not motivated to do much. They rarely feel sudden urges to do things, and so they do not feel the need to be in strict behavioral self-control. They often avoid complication and difficult paths, opting for the simpler path to achieving a goal. At the extreme, these individuals may lack motivation to do anything productive at all.

High N–Low C, the Undercontrolled, represents individuals who have difficulty controlling any of their urges. Even when they know it is not in their best self-interest, these individuals cannot help but satiate their desires. They are impulsive and can be reckless, engaging in self-defeating behaviors. At the extreme, these individuals are likely to be prone to problem behaviors, such as substance abuse, risky sexual behavior, or other behaviors that are potentially unsafe.

High N–High C, the Overcontrolled, represents individuals who feel strong pressure to achieve and tend to exert extreme behavioral control over themselves. These combined tendencies often lead to perfectionism, with very little ability to tolerate even small errors or failures. They become easily shamed and guilty, and at the extreme they can easily become obsessive and/or compulsive.

Low N–High C, the Directed, represents individuals who are able to set goals and attain them, even when they are faced with discouraging setbacks or enticing distractions. They pay little attention to their own impulsive needs, managing to maintain momentum toward achieving their goals at hand. They are not overwhelmed by setbacks and disappointment, pushing forward with their tasks until completion.

Style of Activity

The Style of Activity category relates to the general vitality with which individuals approach life and the tasks that they take on. It is composed of Extraversion and Conscientiousness, which combine to describe both the levels of drive and focus generally found in the individual.

Low E–Low C, the Lethargic, represents individuals who lack enthusiasm and vigor in life and work. They are generally unmotivated to take on challenges, and they can be extremely passive, often taking on followership roles in groups. These individuals do not often come up with the ideas to start new activities or projects, and they can easily get left behind by others who are more highly motivated to achieve.

High E–Low C, the Funlover, represents individuals who are extremely enthusiastic and excitable, but who lack the focus and impulse control to target their energy toward accomplishing goals. They enjoy excitement and novelty in life, often preferring an adventure to an accomplishment. They can be impulsive and hedonistic, sacrificing an ongoing project for a quick thrill.

High E–High C, the Go-Getter, represents individuals who are eager and enthusiastic about accomplishing goals. They work quickly and effectively in a goal-directed

manner, ensuring that they are continually progressing toward their end goals. These individuals are good at identifying what needs to be done, making them good to work with in teams; however, they are unafraid to assert themselves and their work style onto others, which can come across as overbearing. Their eagerness for the task at hand pushes them, and can motivate others, toward successful completion.

Low E–High C, the Plodder, represents individuals who are slow and steady, working consistently on tasks at a sluggish pace but reliably completing what is expected of them. Their rate of working mirrors their rate of being when not working as well, which is unhurried and steady. They approach tasks methodically and rationally, tackling each challenge as it comes. These individuals are good workers and reliable friends and peers, though their slow pace will not suit all tasks and projects.

Style of Attitudes

The Style of Attitudes category relates to individuals' values and beliefs and the general cognitive approach that they use to create and evaluate them. It is composed of Openness and Agreeableness, which help characterize the extent to which individuals are influenced by others and by their own personally held (and past) beliefs.

Low O–Low A, the Resolute Believer, represents individuals who have deeply held personal convictions about human nature and social problems that will be difficult to budge. They are highly moralistic and judgmental, and they believe that everyone should follow rules and laws strictly and should be punished if they do not. These individuals take an unsentimental approach to social problems, favoring discipline and consequences over any solution they would consider sentimental.

High O–Low A, the Free-Thinker, represents individuals who think critically about all options before making a personal judgment about the problem at hand. Uninfluenced by others' opinions, tradition, or emotional appeal, they are open to hearing the rational justification for any side of an issue, and they consider each side as a viable option before coming to a conclusion. They do not necessarily subscribe to others' notion of what is right and wrong, and their decisions about morality may not align with others in their life, which does not bother them at all.

High O–High A, the Progressive, represents individuals who are both rationally open to new ideas and have faith in the inherent goodness of human nature and its potential for improvement through cooperation and innovation. They often believe that education is key to bettering the world, and they are willing to try any and all new ideas to aide humanity. Reason and rationalism are important to these individuals, as they work through many ideas and angles in trying to solve problems.

Low O–High A, the Traditionalist, represents individuals whose values and beliefs stem from their family or context of origin. They believe that the values with which they were raised worked well for a reason and that society (and groups within it) would function better if everyone stuck to them. These individuals do not value questioning the rules or laws of society or coming up with their own form of morality. They prefer to keep peace by following the rules as they were given to them.

Style of Learning

The Style of Learning category relates to the type of scholar and student the individual is likely to be. It is composed of Openness and Conscientiousness, which balance the

openness and eagerness for understanding new things with the focus and rigor to do so methodically.

Low O–Low C, the Reluctant Scholar, represents individuals for whom learning and scholarship are not a priority. They will not self-motivate to learn, study, or do any more work than necessary, and external incentives are often necessary to get them to put in effort. They often have what appear to be attentional or executive functioning problems, having difficulty staying on task, organizing themselves, or planning out their learning effectively. They are not strong intellectually, and scholarly pursuits are not where their interests lie.

High O–Low C, the Dreamer, represents individuals who endeavor to learn new ideas and begin grand new projects, but their follow-through is a weakness. They are often excellent at developing ideas and being creative, but they can easily get lost in their own creative process rather than picking a path and executing it. Unfazed by abstraction, ambiguity, and lack of clarity in an idea or project, these individuals are often more successful when somebody else helps them focus to complete tasks.

High O–High C, the Good Student, represents individuals who both genuinely love learning and have the focus and drive to do it successfully. They are academically and achievement oriented, and they value critically scrutinizing many different angles of problems in order to come up with creative solutions. It is important to note that the NEO does not measure intelligence, so this style does not distinguish more highly intelligent individuals from their peers. However, when combined with intelligence, this style has the potential for individuals to excel at an extremely high level educationally and, ultimately, occupationally.

Low O–High C, the By-the-Booker, represents individuals who diligently follow a plan for learning, going methodically through the steps toward achieving more knowledge. They lack creativity and an appreciation for nuance and ambiguity, strongly prefer concrete answers, and prefer step-by-step methods to reaching learning objectives. Rote learning is a strength, whereas abstracting knowledge to apply it under different conditions or circumstances is often more challenging for them.

Style of Character

The Style of Character category relates to the desire and ability to behave in a way that is either self-focused or other-focused. It is composed of Agreeableness and Conscientiousness, which combine the self versus other orientation with the amount of drive and follow-through that backs it up.

Low A–Low C, the Undistinguished, represents individuals who value their own well-being over that of others but lack the diligence and drive to self-promote effectively. They tend not to be charitable, and they often satisfy their own pleasures in easy-to-obtain ways, which can lead to habits that are unhealthy and hard to break.

High A–Low C, the Well-Intentioned, represents individuals who are genuinely caring about others but who often cannot accomplish their goals to help others effectively. These individuals are better at inspiring others to be kind than accomplishing works of good on their own. Their intentions are genuinely sympathetic toward others, though, despite their difficulty following through with projects to help others.

High A–High C, the Effective Altruist, represents individuals who work tirelessly and effectively to help others, giving themselves fully for the good of the group or

to socially conscious projects. They are able to focus their energy to help others, and they are determined to finish even menial or difficult tasks, if they are for the benefit of others.

Low A–High C, the Self-Promoter, represents individuals who are both primarily interested in and effective at accomplishing their own personal goals. In an individualistic setting or culture, these individuals are often very successful in their chosen fields, as they are determined to reach their own goals and are driven to accomplish them with single-minded focus. They are not deterred from their own successful accomplishments by the suffering or floundering of others. Their ultimate goal is for themselves to succeed, and they will do so even at the expense of others.

RECOMMENDED READING

- Costa, P. T. Jr., & McCrae, R. R. (2009). The five-factor model and the NEO inventories. In J.N. Butcher (Ed.), *Oxford handbook of personality assessment*. New York, NY: Oxford University Press.
- Costa, P. T. Jr., & Piedmont, R. L. (2005). Multivariate assessment: NEO PI–R profiles of Madeline G. In J. S. Wiggins (Ed.), *Paradigms of personality assessment*. New York, NY: Guilford Press.
- McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer Academic/Plenum.
- McCrae, R. R., & Costa, P. T. Jr. (2010). *NEO Inventories: Professional manual*. Lutz, FL: Psychological Assessment Resources.

THE RORSCHACH

The Rorschach is a performance-based test of personality functioning based on interpreting a person's responses to 10 bilaterally symmetrical inkblots. The overall goal of the technique is to assess the structure of personality, with particular emphasis on how individuals construct their experience and the meanings assigned to their perceptual experiences (thematic imagery; Weiner, 2004). The interpretations on Rorschach data can provide information on variables such as motivations, response tendencies, cognitive operations, affectivity, and personal and interpersonal perceptions. Despite attacks from both in and outside the field of psychology, the Rorschach remains one of the most extensively used and thoroughly researched techniques (Archer & Newsom, 2000; Camara, Nathan, & Puente, 2000; C. E. Watkins, Campbell, Nieberding, & Hallmark, 1995). This is reflected in the fact that more than 200 books and 10,000 articles have been written about or using the Rorschach (Exner, 2003).

The central assumption of the Rorschach is that stimuli from the environment are organized by a person's specific needs, motives, and conflicts, as well as by certain perceptual "sets." This need for organization becomes more exaggerated, extensive, and conspicuous when individuals are confronted with ambiguous stimuli, such as inkblots. Thus, they must draw on their personal internal images, ideas, and relationships to create a response. This process requires that persons organize these perceptions as well as associate them with experiences and impressions. The central thesis on which Rorschach interpretation is based is this: The process by which persons organize their responses to the Rorschach is representative of how they confront other ambiguous situations requiring organization and judgment. Once the responses have been made and recorded, they are coded along different dimensions, including the *location*, or the area of the inkblot on which they focused; *determinants*, or specific properties of the blot they used in making their responses (color, shape, etc.); and *content*, or general class of objects to which the response belongs (human, animal, anatomy, etc.). The interpretation of the overall protocol is based on the relative number of responses that fall into each of these categories. Some systems also score for the extent to which subjects organize their responses (organizational activity), the types of verbalizations, and the meaningful associations related to the inkblots.

Although these scoring categories may appear straightforward, the specifics of scoring and interpreting the Rorschach are extremely complex. Furthermore, attempts to develop a precise, universally accepted coding system have not been entirely successful, which creates some confusion and ambiguity in approaching the Rorschach technique itself. Although the primary scoring systems have some agreed-on similarities, there are also significant differences in the elements of these systems. These differences, in turn, reflect the complexity and ambiguity in the nature of the responses made to

the cards. Thus, effective use of the Rorschach depends on a thorough knowledge of a scoring system, clinical experience, and adequate knowledge of personality and psychopathology.

The general purpose of this chapter is to provide an overview of administration, scoring, and interpretation using the two predominant systems currently in use: Exner's Comprehensive System and the Rorschach Performance Assessment System (R-PAS). Exner's system is arguably the most ambitious and widely used Rorschach system to date, and the R-PAS, largely based on the Comprehensive System, is arguably the most psychometrically supported Rorschach system to date. Both systems include some of the most frequently used scorings and interpretations from other systems.

Scoring for the Comprehensive System and R-PAS is quite complex, and only a brief overview can be covered in this chapter. Clinicians who wish to use precise scoring tables and criteria, as well as more extensive elaborations on interpretation, are encouraged to consult Exner and his colleagues' original works (Exner, 2000, 2001, 2003; Exner & Weiner, 1995) for the Comprehensive System and the coding, scoring, and interpretive manual for the R-PAS (Meyer et al., 2011). This chapter cannot stand as a substitute for the depth of these works. Its major intent is to familiarize persons with the Rorschach in general and, more specifically, with Exner's and the R-PAS's approach to interpretation. In addition to students who are learning the systems, persons who are already familiar with Exner's system or the R-PAS might wish to consult sections of this chapter to obtain summaries of different scoring categories and interpretive hypotheses. This might be most appropriate for practitioners who use the Rorschach only occasionally. Finally, persons who use other scoring systems may wish to consult the different interpretive hypotheses as an aid to interpretation. Doing this is theoretically possible because Exner and the R-PAS both incorporated the major approaches from other systems into their systems. However, variations are likely to occur among the Comprehensive System, R-PAS, and other systems; therefore, interpretations should be made with caution.

HISTORY AND DEVELOPMENT

Many inkblot-type tests and games had existed long before Rorschach published his original 10 cards in 1921. For example, da Vinci and Botticelli were interested in determining how a person's interpretations of ambiguous designs reflected his or her personality. This theme was later considered by Binet and Henri in 1895 and by Whipple in 1910. A popular parlor game named Blotto that developed in the late 1800s required players to make creative responses to inkblots. However, Rorschach developed the first extensive, empirically based system to score and interpret responses to a standardized set of cards. Unfortunately, Rorschach died at age 37, shortly after the publication of his major work, *Psychodiagnostik* (1921/1941). His work was continued to a limited extent by three of his colleagues—Emil Oberholzer, George Roeurer, and Walter Morgenthaler.

The main approach used by Rorschach and other early developers of inkblot techniques was to note the characteristic responses of different types of populations. Thus, the initial norms were developed to help differentiate among various clinical

and normal populations: individuals with schizophrenia, persons with intellectual disabilities (mental retardation), normals, artists, scholars, and other subgroups with known characteristics. Rorschach primarily wanted to establish empirically based discriminations among different groups and was only minimally concerned with the symbolical interpretation of content. Many of his original concepts and scoring categories have been continued within current systems of analysis. For example, he noted that depressed, sullen patients seemed to give the fewest responses. Persons giving a large number of very quick responses were likely to be similarly “scattered” in their perception and ideation in nontest situations. He also considered the importance of long latencies (so-called shock responses) and hypothesized that they were related to a sense of helplessness and emotional repression.

Had Rorschach lived longer, the history and development of his test might have been quite different. Without the continued guidance and research from the “founding father,” the strands of the Rorschach technique were taken up by persons who had quite different backgrounds from Rorschach and from one another. By 1957, five Rorschach systems were in wide use, the most popular being those developed by Beck and by Klopfer. These two approaches came to represent polarized schools of thought and were often in conflict.

S. J. Beck (1937) adhered closely to Rorschach’s format for coding and scoring. He continually stressed the importance of establishing strong empirical relationships between Rorschach codes and outside criterion measures. Beck emphasized that the response to the Rorschach involved primarily a perceptual-cognitive process in which the respondents structure and organize their perceptions into meaningful responses. This perceptual-cognitive process was likely to reflect their responses to their world in general. For example, persons who broke down their perceptions of an inkblot into small details were likely to behave similarly for perceptions outside the testing situation.

In contrast, B. Klopfer (1937) was closely aligned to phenomenology and the theories of personality developed by Freud and Jung. As a result, he emphasized the symbolical and experiential nature of a respondent’s Rorschach contents. Thus, Klopfer believed that Rorschach responses were fantasy products triggered by the stimulus of the inkblots. For example, persons who perceived threatening objects on the inkblots would perceive aspects of their world as similarly threatening. Although not as popular, additional systems developed by Piotrowski, Hertz, and Rapaport represented a middle ground between the two extremes represented by Beck and Klopfer.

With five distinct systems available, the Rorschach became not a unitary test but five different tests. Exner (1969) provided a comparative analysis of these different systems and later concluded that “the notion of *the* Rorschach was more myth than reality” (Exner, 1986, p. 19). He pointed out that none of the five systems used the same verbal instructions and only two of the systems required identical seating arrangements. More important, each systematizer developed his or her own format for coding, which resulted in many differences regarding interpretation, the components required to calculate quantitative formulas, the meanings associated with many of the variables, and the interpretive postulates.

The wide range of often-competing approaches resulted in numerous detrimental practices. A survey of practitioners by Exner and Exner (1972) indicated that 22% of all

respondents had abandoned scoring altogether and instead based their interpretations on a subjective analysis of content. Of those who did score, 75% used their own personalized integration of scores from a variety of systems. In addition, the vast majority did not follow any prescribed set of instructions for administration. With researchers using a variety of approaches, comparison of the results of different studies was difficult. Researchers in the early 1970s further reported difficulties in recruiting subjects, problems with experimenter bias that needed to be corrected by using multiple examiners, statistical complexities of data analysis, inadequate control groups, and insufficient normative data (Exner, 1993, 2003). Some of the elements had no empirical basis. The general conclusion, based on these findings, was that the research on and the clinical use of the Rorschach were seriously flawed. Despite this, all five systems included some empirically sturdy elements.

To correct the difficulties with both the research and clinical use of the Rorschach, Exner and his colleagues began the collection of a broad normative database and the development of an integrated system of scoring and interpretation. Their initial step was to establish clear guidelines for seating, verbal instructions, recording, and inquiry by the examiner regarding the examinee's responses. The best features for scoring and interpretation, based on both empirical validation and commonality across systems, were adapted from each of the five different systems. A scoring category was included in the new system only after it had achieved a minimum .85 level for interscorer reliability. The final product was first published in 1974 as *The Rorschach: A Comprehensive System* and has since been released in second (Exner, 1986), third (Exner, 1993), and fourth (Exner, 2003) editions. A second volume relating to current research and interpretation has been released in two editions (Exner, 1978, 1991), and two editions of a volume on the assessment of children and adolescents have also been published (Exner & Weiner, 1982, 1995).

Normative data for the Comprehensive System has undergone continual revision. A major reason for these revisions has been to refine stratification. A further impetus was that in 1990, the Comprehensive System eliminated all protocols with fewer than 14 responses because these were likely to have resulted in invalid protocols. The normative base reported in Exner's 1993 (third) edition of the Comprehensive System was composed of 700 adult nonpatients and 1,390 nonpatient children and adolescents between the ages of 5 and 16. However, it was discovered in 1999 that more than 200 duplicate adult protocols had inadvertently been included. As a result, a new normative sample was begun. The most recent publication has included 450 contemporary protocols from persons aged 18 to 65+, evenly divided between males and females, with a wide range of education and a variety of ethnic groups (Exner & Erdberg, 2005). Future publications will include a progressively larger number of participants. In addition, an international normative reference group has been collected by Meyer, Erdberg, and Shaffer (2007). The child and adolescent sample reported in Exner (2003) is the same as that included in 1993 (includes 1,390 nonpatients between the ages of 5 and 16).

Exner's integration of the different Rorschach approaches into his Comprehensive System has been successful in that most research studies over the past 20 years have used his system, and it has become by far the most frequently taught system in graduate training. His attention to empirical validation, combined with a large normative database, has served to increase its acceptance and status. Access to training and

interpretive aids has been facilitated through numerous workshops, a scoring workbook (Exner, 2001), ongoing research publications, new editions of earlier volumes, and computer-assisted scoring and interpretation (Exner, 1984, 1986, 1993, 2003).

Debates regarding the psychometric adequacy of the Rorschach have created one of the greatest controversies in the history of psychology. From the beginning, the Rorschach was met with skepticism in the United States; yet it developed a strong following. At one point, the Rorschach was the second most frequently used test, and, in the 1940s and 1950s, the name *Rorschach* was almost synonymous with clinical psychology. Despite this initial (and continuing) popularity, reviews have generally been quite critical. As early as 1954, Shaffer declared that the Rorschach could no longer be considered a promising instrument. Eleven years later, Dana (1965) somewhat prematurely concluded: "Indeed, we have come to the end of an era, preoccupation with the Rorschach as a test" (p. 495). A. R. Jensen (1965) was even more critical when he recommended that "the Rorschach be altogether abandoned in clinical practice, and that students in clinical psychology not be required to waste their time learning the technique" (p. 509). While Garb (1999) called for a "moratorium" on its use until research clarified which scoring categories are valid, his response to the Mihura, Meyer, Dumitrascu, and Bombel. (2013) meta-analysis reversed this decision, at least partially (Wood, Garb, Nezworski, Lilienfeld, & Duke, 2015).

It should be noted that one of the early difficulties in establishing the psychometric properties of the Rorschach was in making meaningful comparisons across various studies. As Exner (1969, 1974, 1986, 1993, 2003) has repeatedly pointed out, there is not *a* Rorschach; rather, at least five different Rorschachs have been created around the five major systems. Reliability and validity studies performed on one system did not necessarily mean that the findings from these studies could be generalized to any of the other systems. However, reviewers often acted as if there were only one Rorschach. Furthermore, many studies were poorly conducted. They were characterized by inadequate controls for age, sex, race, IQ, and socioeconomic status. In addition, many studies had extremely wide variations in the training required for scorers, insufficient protection from experimenter bias, poor validation criteria, and inadequate statistical models. These difficulties were amply demonstrated when Exner (1986) and his associates found it necessary to discard 1,400 research studies of a total of 2,100 studies published before 1970.

More recently, the depth and sophistication of the criticisms have increased. This has resulted in extensive arguments and counterarguments, with each side citing numerous studies in favor of their positions. Between 1998 and 2003, most major assessment journals published special series debating the relative merits of the Rorschach. Challenges were directed at nearly all aspects of the test, including the adequacy of its norms, interscorer reliability, temporal stability, the accuracy of meta-analysis that had found support for the Rorschach, and its level of incremental validity. The central elements of these debates are integrated in this chapter into the sections titled "Reliability and Validity" and "Assets and Limitations."

Over the past five decades, Exner was responsible for much of the leadership and many of the advances regarding the Rorschach. In 1997, Exner established the Rorschach Research Council (RRC; Exner, 1997), which was primarily charged with pushing forward the research agenda and empirical base of the Comprehensive

System. Upon Exner's death in 2006, however, there was a lack of clarity around how the Comprehensive System could evolve and who would continue its development. The result was a group of researchers, deeply entrenched in the Comprehensive System and its research (and many from the RRC), developing a new scoring system based on the best available published research on the Rorschach (largely from the Comprehensive System, but incorporating other promising variables as well). This group developed the Rorschach Performance Assessment System (R-PAS).

The development of the R-PAS had six stated, explicit goals (Meyer, Viglione, Mihura, Erard, & Erdberg, 2011). These goals significantly strengthen the test and respond to a number of criticisms (Groth-Marnat, 2009; Meyer et al., 2007). These goals included eliminating variables with limited empirical support and including variables with strong empirical and clinical support, including a description of the empirical basis of interpretations of variables throughout; basing the system on international norms; simplifying the system itself; statistically adjusting variable scores based on the overall complexity of the protocol; "optimizing" the number of responses given by respondents by introducing procedures to restrict the range of number of responses (while ensuring that these procedures do not decrease interpretability of the test); and developing new, statistically derived indices. Additionally, the R-PAS provides a web-based, secure scoring program. The resulting system has led even the most vocal critics of the Rorschach (Garb, Lilienfeld, and colleagues) to endorse the use of at least parts of the test in clinical and research settings (Wood et al., 2015). The bulk of variables these critics endorsed mirror the structure of the R-PAS.

RELIABILITY AND VALIDITY

As noted previously, Exner originally included only scoring categories that had interscorer reliabilities of .85 or higher. Some controversy has resulted concerning these values in that other researchers have reported greater variability. Parker (1983) analyzed 39 papers using 530 different statistical procedures published in the *Journal of Personality Assessment* between 1971 and 1980. He concluded that, overall, the Rorschach can be expected to have reliabilities in the low to middle .80s. However, only two of his studies used the Comprehensive System. Acklin, McDowell, Verschell, and Chan (2000) found that nearly half of the categories for the Comprehensive System showed excellent reliabilities ($>.81$) with substantial reliability (.61–.80) for a third of the categories. They concluded that a majority of the categories had excellent interscorer reliability, but a subset of about a quarter of the variables demonstrated less than adequate ($<.61$) reliability. The problem with the Acklin et al. data, however, was that the sample sizes were small, with the result that greater variability would be expected.

In the most ambitious, rigorous, and large-scale study to date, Meyer et al. (2002) used eight different data sets and employed several different strategies to determine the reliability of the categories for the Comprehensive System. They concluded that it had overall excellent interscorer reliabilities with median correlations ranging from .82 to .97, depending on the data set used. Exner (2003) reported new interscorer reliabilities with agreement ranging from a high of 99% for texture and vista responses to a

low of 88% for passive movement. These correlations support the claims of Exner and of Gronnerod (2006) that, if scorers are appropriately trained, the system has excellent interscorer reliabilities. However, some research has demonstrated that interscorer reliability was not as good for codes that occur infrequently (Acklin, McDowell, Verschell, & Chan, 2000; Meyer et al., 2002; Viglione & Taylor, 2003). Interscorer reliability for the R-PAS, much newer and thus not as heavily tested, appears to be good to excellent (Viglione, Blume-Marcovici, Miller, Giromini, & Meyer, 2012), though similar to the Comprehensive System. As expected, the lowest reliabilities tend to accompany codes that appear infrequently.

An additional crucial area for reliability is the extent to which clinicians agree on interpretations related to test data. If one clinician made interpretations that were at variance with those of other clinicians, it would not only indicate low inter-interpretor reliability, but some of the interpretations would necessarily be inaccurate. However, Meyer, Mihura, and Smith (2005) found that interpretive agreement on the Comprehensive System among experienced clinicians ranged between .76 and .89, which is quite good.

Test-retest reliabilities for the Comprehensive System have been somewhat variable. Retesting of 41 variables over a 1-year interval for a nonpatient group produced reliabilities ranging between .26 and .92 (see Table 11.3 in Exner, 2003, p. 179). Four of the correlations were above .90, 25 were between .81 and .89, and 10 were below .75. Exner (2003) clarified that the 10 variables below .75 would all be expected to have had relatively low reliabilities because they related to changeable state (rather than trait) characteristics of the person. He also pointed out that the most important elements in interpretation are the ratios and percentages, all of which were among the higher reliabilities. Retesting for the same group over a 3-year interval produced a similar but slightly lower pattern of reliability. In contrast, another group of nonpatient adults, retested over a much shorter (3-week) interval, had somewhat higher overall reliabilities than for either the 1-year or 3-year retestings (Exner, 1986). A more extensive summary of test-retest reliability by Meyer and Archer (2001) found that the mean reliability was .66 (range from .46 to .84, *Mdn* = .69). This is similar to the .66 to .82 mean reliabilities summarized by Viglione and Hilsenroth (2001) and the .67 for a quite long retesting of 5 years by Gronnerod (2006). Gronnerod (2004) also found that many of the elements of the Rorschach were valid indicators of change following psychotherapy. One issue, however, is that the Comprehensive System has approximately 125 variables, and some of these do not have known test-retest reliability on them. The number of these untested reliabilities varies across researchers, with Wood and Lilienfeld (1999) stating that 85 variables have missing reliabilities and Viglione and Hilsenroth (2001) stating that only 12 variables have unknown test-retest reliabilities. Test-retest reliability has yet to be examined with the R-PAS specifically, though it would be expected to be similar to the Comprehensive System.

Long-term Comprehensive System retesting for children has not come close to the same degree of stability as for adults (Exner, 2003; Exner & Weiner, 1995). Exner (1986) clarified that this low stability for test results is to be expected, given that children undergo considerable developmental changes. However, short-term retesting over 7-day (for 8-year-olds) and 3-week (for 9-year-olds) intervals did indicate acceptable levels of stability (Exner, 2003). Only 2 of 25 variables were below .70, with at least 7

above .90 and the remainder from .70 to .90. As with adults, the ratios and percentages demonstrated relatively high stabilities. Although acceptable short-term stability for young children's Rorschach variables was demonstrated, long-term stability was not found to occur until children reached the age of 14 or older (Exner, 2003; Exner, Thomas, & Mason, 1985).

The primary focus of early validity studies was to empirically discriminate among different populations. These empirically based discriminations were originally based on past observations of a particular group's responses to the Rorschach, the development of norms based on these responses, and comparisons of an individual's Rorschach responses with these norms. For example, a person with schizophrenia might have a relatively high number of poor-quality responses, or a person with depression might have very few human movement responses. In addition to these empirical discriminations, efforts have been made to develop a conceptual basis for specific responses or response patterns. Thus, it has been conceptualized that people with schizophrenia have poor-quality responses because they do not perceive the world the way most people do; their perceptions are distorted and inaccurate, and their reality-testing is poor. A further approach, which was not extensively developed in the Comprehensive System (nor by Rorschach himself), was the validation of the latent meaning of symbolical content.

These very general approaches have given rise to a surprisingly large number of specific codes, scores, and interpretations, all of which have had various degrees of support. Many of the early validity studies are difficult to evaluate because of the varying scoring systems and poor methodologies. In addition, most early studies depended on inadequate norms (especially for studies conducted on children, adolescents, cross-cultural groups, and persons over 70). Test results might also have been significantly influenced by situational and interpersonal variables, such as seating, instructions, rapport, gender, and personality of the examiner (see review by Masling, 1992). It should then come as no surprise that, for every study supporting an interpretive hypothesis, there would often be another refuting the same hypothesis.

Some efforts have been made to look at the Comprehensive System as a whole to evaluate its overall validity for its intended purposes. Early meta-analyses indicated that validity ranged from .40 to .50 (Atkinson, Quarington, Alp, & Cyr, 1986; Parker, 1983; Parker, Hanson, & Hunsley, 1988; Weiner, 1996). However, these results have been challenged by Garb, Florio, and Grove (1998; Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Hunsley & Bailey, 2001), who reanalyzed the data from Parker et al. and concluded that the overall validity coefficients for the Rorschach were only .29 (in contrast to the significantly higher validity of .48 for the Minnesota Multiphasic Personality Inventory [MMPI]). This finding produced lively debates in the literature regarding the most appropriate methods of analysis. The majority of recent meta-analyses have continued to support the validity of the Rorschach (R. F. Bornstein, 1999; Meyer, 2004; Meyer & Archer, 2001; Meyer & Handler, 1997; Meyer et al., 2005). However, interactions with type of scoring system, experience of the scorer, and type of population used were likely to have complicated the picture. This approach to evaluating global validity for an extremely multifaceted measure is limited, as it does not take into account the validity of individual scales and indexes (Lilienfeld,

Wood, & Garb, 2000), each which measure very different aspects of the individual. As such, it is much more important to establish validity of the individual variables.

Establishing the validity of Rorschach variables has been complicated by the many scoring categories and quantitative formulas, each of which has varying levels of validity. When narrowed down to the Comprehensive System, the test manual itself (Exner, 2003) provides empirical citation for the validity of all of the variables used within it. However, about 100 of the studies cited in the manual are unpublished studies conducted by Exner himself (Lilienfeld et al., 2000; Wood, Nezworski, & Stejskal, 1996), which raises significant questions about the integrity of the test. A further area of difficulty in establishing validity is that Exner cited extensive validity studies throughout his three volumes, but many of these studies were not done using his Comprehensive System. Comparability between the different studies and systems is frequently assumed or at least implied. However, often these studies were performed at a time when norms were inadequate, interscorer reliability was questionable, and little concern was given to the possible confounding effects of age, intellectual level, education, and verbal aptitude. The development of the Comprehensive System itself was largely motivated by the deficiencies (and strengths) inherent in each of the earlier systems. As such, a great deal of further effort has been spent on evaluating the validity of individual variables within the Comprehensive System.

Some interpretations have greater validity than others even in a specific category. For example, the number of human movement responses (M) has been used as an index of both creativity and fantasy. A review of the research by Exner (1993) indicated that M relates fairly clearly to fantasy in that it has been correlated with daydreaming, sleep/dream deprivation, dream recall, and total time spent dreaming, whereas associations between M and creativity have been weaker and more controversial. Validity might also depend on the context and population for which the test is used. For example, a Comprehensive System Depression Index (DEPI) based on seven Rorschach combinations of scores has been found to provide low or no associations with the presence of depression among adults (Jorgensen, Anderson, & Dam, 2000; Meyer, 2000; Mihura et al., 2013). Among adolescent populations, the DEPI was not successful in distinguishing those with depression from those with schizophrenia (Archer & Krishnamurthy, 1997a; Ball, Archer, Gordon, & French, 1991; Stredny & Ball, 2005). In contrast to DEPI, a Comprehensive System index designed to detect thought disorders (Perceptual Thinking Index [PTI]) has been quite successful (Dao & Prevatt, 2006; Mihura et al., 2013). Mihura and her colleagues (2013) published an extremely important meta-analysis of the validity studies of individual Comprehensive System variables; it was largely on the basis of this information that variables were included and excluded from the R-PAS. Additional validity data on specific scoring categories and formulas are included in the "Interpretation" section of this chapter. These data should be carefully read to more fully understand Rorschach validity.

One of the main efforts toward establishing Rorschach validity has been directed toward determining its ability to discriminate among different types of populations. The success of these differentiations has been somewhat equivocal (see Vincent & Harman, 1991; Wood, Lilienfeld, Garb, & Nezworski, 2000). For example, Wood et al. (2000) have indicated that, with the exception of a few disorders (schizophrenia,

borderline personality disorder, bipolar disorder), the Rorschach has not been very effective at assisting with making formal psychiatric diagnoses. A defense is that, in contrast to structured interviews or tests such as the MCMI, the Rorschach was not designed to accomplish this goal. As the Rorschach was not designed to provide diagnoses, diagnostic accuracy is likely to be accurate for some diagnoses but not for others. Although the Rorschach is not the optimal instrument for most forms of diagnosis, it has been found to effectively predict variables such as outcome from psychotherapy (using the Prognostic Rating Scale; $r = .45$), detection of psychosis (using the Schizophrenia Index; $r = .44$), and dependent behavior (using the Oral Dependency Scale; $r = .37$; Meyer & Archer, 2001).

One major factor that may serve to lower Comprehensive System validity is the meaning associated with, and the effects of, response productivity. Various interpretations have been associated with extremes of productivity, with low productivity suggesting defensiveness, depression, and malingering and extremely high productivity suggesting high achievement or an obsessive-compulsive personality. However, response productivity has also been found to be closely tied to age, intellectual level, verbal aptitude, and amount of education. Norms have been provided for different ages (Exner, 1993; Exner & Weiner, 1995), which can be helpful in correcting for the effects of age. However, intellectual level, verbal aptitude, and amount of education can potentially confound the meanings associated with response productivity. A high number of responses does not necessarily represent traditional personality interpretations (obsessiveness, creativity, good impulse control) but might merely indicate a high level of verbal aptitude.

Most early validity studies rarely considered the preceding factors. More important, the number of responses not only affects interpretations related specifically to response productivity; productivity also affects many other areas of interpretation. For example, a low number of responses is likely to increase the relative number of responses based on the whole inkblot (W). In contrast, a high number of responses would be likely to increase the relative number of small detail (Dd) responses. Because interpretations are frequently based on the relative proportions of different scoring categories (calculated in quantitative formulas), the overall number of responses is likely to influence and possibly compromise the validity of the formulas. However, Exner (1993) found that lengthy records generally did not result in different interpretations when compared with records from the same persons with average numbers of responses. For practical reasons, he has recommended that the number of responses be limited if the person gives six or more responses to the first card or five or more responses to the second card (see the "Administration" section of this chapter). In contrast to lengthy protocols are ones with extremely low numbers of responses. Exner (2003) recommended that brief protocols (fewer than 14) be discarded and the test be readministered. This problem with the meaning of various numbers of responses largely led Holtzman to develop his alternate test (Holtzman Inkblot Test), in which subjects provide only one response for each inkblot in his series (Holtzman, 1988). The R-PAS restricts the possible range of number of responses (by prompting for more or moving on to the next card at different particular moments in administration), which can help correct for response productivity affecting the test variables without altering their interpretation.

A significant concern with the Comprehensive System is that scores on the test indicate more pathology when compared with indicators from other sources (Hamel, Shaffer, & Erdberg, 2000; Shaffer, Erdberg, & Harioan, 1999). If true, this finding would indicate significant concern related to child custody and other forensic and clinical decisions that might be based on Comprehensive System data. Advocates of the Rorschach have replied that overpathologizing may appear to be present in part because the Rorschach norms were based on persons who were not merely nonpatients but were rated as being healthy and well functioning (Meyer, 2001). The basis for these norms would make it fairly easy for many people to appear relatively pathological compared to them. However, Rorschach norms, research, and decision rules are being continually revised and improved, specifically now with the onset of the R-PAS. One implication of this controversy is that Comprehensive System scores (and inferences based on them) should be checked against other sources of information. If this information does not support the Comprehensive System, then the Rorschach inferences should be treated with considerable skepticism. The R-PAS chose to use the international norms (Meyer et al., 2007) in part to correct for the criticism of overpathologizing, as the international norms represent a clearer normative reference for nonpatients. Use of these norms is encouraged in general, even with the Comprehensive System, and any cross-cultural use of the Rorschach should use the international normative reference group.

This overview of Rorschach reliability and validity suggests a number of conclusions. Interscorer and test-retest reliabilities for the Comprehensive System have generally been supported, although there are a number of variables with unknown test-retest reliability. The R-PAS also evidences good interscorer reliability in general, though test-retest reliabilities have not been studied well yet. The overall validity of the Comprehensive System has been found to be moderate (.30–.50), though this work has been met with criticism. The Mihura et al. (2013) meta-analysis made a big contribution in the move toward establishing more credible estimates of validity of individual variables. Overall, across all the studied variables (which included the majority of variables in the Comprehensive System), they found an average medium effect size of .27 (which is expected and acceptable for a personality measure), but individual variables varied widely as to their evidence base. As a result, the R-PAS retained only variables that had adequate effect sizes in the hypothesized directions. It is strongly recommended that even when using the Comprehensive System, practitioners review the meta-analysis to understand which variables should be interpreted with confidence and which variables the practitioner should consider not using.

ASSETS AND LIMITATIONS

As mentioned, the Rorschach has been surrounded by controversy. Often, battle lines have been polarized into either “clinical loyalists” or “academic iconoclasts” (Parker, 1983). Despite thousands of research studies, these positions have changed only minimally over the past 60 years. Masling (2006) suggested that the controversial status of the Rorschach may be largely the result of researcher bias in selectively processing

the voluminous research. It is hoped that with further research on the Comprehensive System and the development of the R-PAS, along with reviews such as those of Meyer and Archer (2001) and Mihura et al. (2013), there will eventually emerge a middle ground that will satisfy hard-nosed empiricists and address areas relevant to clinicians.

Part of the reason the Rorschach has continued to have such high popularity is the number of attractive features associated with it. Perhaps part of its allure is the mystery it frequently invokes. How could something as seemingly simple as 10 inkblots reveal inner aspects of a person's personality? Metaphors such as "X rays of the mind" have certainly served to enhance its mystery and power. Often a Rorschach protocol is perceived as something like a deep well into which a skilled clinician can dip repeatedly, continually coming up with rich and valuable information. The practitioner is framed as a seer and an artist rather than a technician. Indeed, there are many anecdotes in which highly trained Rorschach experts have provided in-depth, nuanced descriptions of a wide range of client characteristics.

One frequently noted asset is that the Rorschach is considered to be excellent at bypassing a person's conscious resistance; instead, it assesses a person's underlying, unconscious structure of personality. Whereas cognitive assessment would never rely entirely on self-report (a self-report rating scale for one's own IQ would never be considered), personality assessment has become dominated by self-report. Performance-based measures of personality, like performance-based measures of cognitive functioning, add an "objective" component to an assessment that is not susceptible to the biases or blind spots, or even motivated deception, of the respondent. This asset might be particularly important if a person appears to have an adequate surface level of adjustment yet the clinician suspects there may be some underlying pathology. Even a structured, self-report test like the MMPI may have difficulty assessing these more hidden levels of pathology. It is precisely the difficulty in organizing the ambiguous Rorschach stimuli that is likely to bring out these latent levels of pathology. There is some support for this view in that persons with lower-level borderline psychopathology have relatively normal performance on structured tests. In contrast, they tend to show clear indications of thought disorder on the far less structured Rorschach (Edell, 1987). Similarly, a relatively hidden trait such as alexithymia has been found in psychosomatic patients based on their Rorschach responses (Acklin & Bernat, 1987). G. Frank (1990) reviewed the existing literature and found that the Rorschach was sensitive to underlying schizophrenic processes even before their clinical expression.

A related asset is the Rorschach's purported high resistance to faking. It is argued that, because the true meanings of the Rorschach responses are not easily understood by untrained individuals, the respondent cannot easily invent faked responses. Some proponents have even stated that it is virtually impossible to fake a Rorschach. Like many other statements about the Rorschach, this one has become quite controversial. Exner (1993, 2003) presented material, from a theoretical and empirical perspective, suggesting that persons developing a Rorschach response go through a series of six stages, one of which is censorship. Respondents seem to come up with far more responses than they present to the examiner, and they select the ones they feel are most appropriate to reveal. Persons who feel emotionally close to the examiner tend to provide more responses and conceal less (Leura & Exner, 1978). This finding raises the possibility that such persons might also have enough control over their responses to

effectively fake a protocol. Thus, responses might depend to a certain extent on social desirability, perceptual accuracy, the context of the assessment, and personal needs.

Despite the possibility of censorship, which might potentially lead to undetected faking, Exner and Wylie (1975) reported that only 1 student in 12 could simulate a profile of an individual with schizophrenia, even though the students were familiar with protocols from actual people with schizophrenia. Specifically, malingerers were likely to have longer free associations (presumably because they were censoring and elaborating on their responses), relatively accurate perceptions, and highly dramatic and idiosyncratic responses (e.g., "That's too awful to look at"). Similarly, Frueh and Kinder (1994) found that persons who were malingering with posttraumatic stress disorder provided responses that were overly dramatic, relatively unrestrained, and indicative of an exaggerated sense of impaired reality testing. Finally, L. S. Grossman, Wasylw, Benn, and Gyoerkoe (2002) noted that sex offenders who minimized psychopathology on the MMPI-2 still had Rorschach protocols that indicated psychopathology.

In contrast to this research, Albert, Fox, and Kahn (1980) found that Rorschach experts did poorly when requested to blindly classify protocols from normals who were requested to fake paranoid schizophrenia, normals taking a standard administration, and those diagnosed with paranoid schizophrenia. Likewise, computer analyses of the same protocols were unsuccessful in effectively detecting faking (M. W. Kahn, Fox, & Rhode, 1988). Although this finding clearly challenges the unfakability of the Rorschach, the Albert et al. and Kahn et al. studies did not simulate the manner in which the Rorschach is likely to be used in clinical practice. Typically, practitioners have knowledge regarding the history of the person, context of the assessment, and behavioral observations, all of which potentially sensitize them to the possibility that a protocol might be faked. Consistent with this was the Frueh and Kinder (1994) study, which found that relevant behavioral observations were at least as important in detecting malingering as the actual scored protocols.

One clear asset of the Rorschach is its ease of administration. The cards can be easily handled, and the total administration time (including inquiry) is typically 50 minutes (Ball, Archer, & Imhof, 1994) for the Comprehensive System and is likely to be shorter for the R-PAS, because of targeted administration strategies. In contrast to the relative ease of administration, coding, scoring, and interpretation are often quite complicated and time-consuming. Clinicians report that scoring usually takes 45 minutes and interpretation requires 50 minutes more (Ball et al., 1994). This means that, collectively, the entire procedure takes nearly 2.5 hours. However, computer-assisted scoring and interpretation is expected to reduce significantly the time for both scoring and interpretation.

Besides the advantages associated with the Rorschach, it has a number of limitations. Although both reliability and validity have generally reached adequate levels, validity is often quite variable across different coding and scoring categories and formulas; some have quite good validity whereas others are moderate, controversial, or even nonexistent (see Mihura et al., 2013). It is usually difficult for the average user to appreciate and take into account the disparate levels of validity when actually making his or her interpretations on the Comprehensive System. It is the responsibility of practitioners to fully understand the validity of each individual variable on the instrument

when using it; the development of the R-PAS has simplified this process by basing inclusion of variables on established validity criteria.

Because the Rorschach is one of the most complex psychological tests in current use, error can potentially be introduced from many different directions, including censorship by the client, administration and coding errors (particularly for infrequently used codes), poor handling of the subtleties of interpretation, incorrect incorporation of the implications of age or education, or possible examiner bias (illusory correlation, primacy effects, etc.). One temptation is to reduce the complexity of the data by using a single-sign approach rather than viewing each sign in the context of the overall configuration. Rorschach “elevations” are often subject to a number of possible interpretive hypotheses, so a single-sign approach is particularly open to error. Thus, interpretations must be continually checked and rechecked against the overall Rorschach configuration, additional test data, and the client’s history.

The complexity of the Rorschach also requires that potential users undergo extensive training. Each new coding category and index that is introduced may add to this problem. In the past, graduate schools sometimes provided a full-semester course on the Rorschach. Some authors, feeling that this is insufficient, have stated that the optimum amount of time is two full-semester courses devoted exclusively to the Rorschach (Hilsenroth & Handler, 1995), a curriculum that is difficult for many programs to justify for two reasons. First, many other tests are both more time efficient and are believed to have superior psychometric properties. Second, the past 25 years have brought a significant increase in the roles and skills required of graduate students, including skills in the area of assessment (neuropsychology, behavioral assessment), as well as in other areas of clinical practice (family therapy, rehabilitation, new modes of intervention, treatment of chronic pain, etc.). Training on the Comprehensive System and/or R-PAS is extremely complicated and time consuming, and many training programs have limited time and resources to use on a single instrument like the Rorschach.

The Rorschach has often been considered to have limited use with children, particularly those under the age of 14 years (Klein, 1986). Reliabilities have been found to be adequate for short-term assessments but clearly inadequate over a long-term basis. Thus, for purposes such as child custody decisions, where longer-term predictions are required, the Rorschach would be quite limited. Any use of the Rorschach for children should make clear that descriptions are only for the short term.

A final consideration, which has implications for both research and practice, is that the large number of variables is likely to produce spurious random significance (Karson, 2005). Wechsler subtest interpretation has attempted to correct for this possibility by carefully calculating the significance of subtest differences, including correction factors for the number and reliabilities of variables considered (see Chapter 5). In contrast, it is difficult to know when the numerous variables considered in the Rorschach might indicate “significance” simply because of random fluctuations of scores (e.g., a .05 level of significance would mean that “significance” would happen by chance in 1 of 20 variables considered). Rorschach interpreters must, therefore, take extra caution with their interpretations.

In summary, the Rorschach is difficult to evaluate because of its complexity, the frequent controversy surrounding it, and considerable variability related to the validity

of its variables. The voluminous research associated with the Rorschach is often both an asset and a limitation. Sorting through the maze of sometimes contradictory findings is difficult. Directing this wealth of research toward a clear understanding of the interpretive meanings associated with certain patterns of scores is especially difficult. Mihura and colleagues (2013) did the field a great service by organizing the wealth of research on the validity of individual variables into a single publication, and the development of the R-PAS based on this and other validity data represents a step forward for the test. The specific assets of the Rorschach are potential wealth of information, simplicity of handling, ability to bypass conscious resistance, and possible resistance to faking. Significant weaknesses are moderate and sometimes quite variable reliabilities and validities (especially when using the Comprehensive System); time required for coding, scoring, and interpretation; limited use with children; extensive time required for training, and possible introduction of error, especially spurious random significance as a result of the large number of areas considered.

USE WITH DIVERSE GROUPS

It has been argued that since the Rorschach is a nonverbal, performance-based measure, it is relatively culture free and therefore ideally suited to assessing ethnic and cross-national populations (Allen & Dana, 2004; Dana, 2005). There is some support for this argument. Specifically, Meyer (2002) and Meyer et al. (2007) found that there was no evidence for the differential validity of the Rorschach Comprehensive System among various ethnic groups. Similarly, Presley, Smith, Hilsenroth, and Exner (2001) compared 23 core scores for African Americans with a matched group of White Americans and found that only one score was significantly different (African Americans produced fewer cooperative movement responses). Meyer, Giromini, Viglione, Reese, and Mihura (2015) found no reliable differences on the R-PAS based on gender, ethnicity, or adult age. Finally, Meyer et al. (2007) found that there were few major differences among an international set of norms on adults derived from 17 different countries. In contrast, he noted that there was far more variability among children's and adolescent's scores. As a result, he discouraged clinicians from making interpretations of child and adolescent psychopathology. However, some newer research has supported the use of the R-PAS and international norms with adolescents (Reese, Viglione, Giromini, 2014; Tibon Czopp, Rothschild-Yakar, & Appel, 2012).

These findings indicate that clinicians working in cross-national settings should use the international norms published by Meyer et al. (2007), and even those working within the United States should consider using them. In addition, any interpretations should also take into account the cultural context. For example, cultures give different value to introversion versus extraversion. Thus, interpretations should take this into account. In addition, the symbolical value related to qualitative interpretations should be particularly sensitive to various cultural meanings. The client's relative degree of acculturation should also be considered. Finally, clinicians should be cautious when interpreting the R-PAS with child and adolescent populations until more research is conducted.

COMPREHENSIVE SYSTEM: ADMINISTRATION

Examiners should standardize their administration procedures as much as possible. This is particularly important because research has consistently indicated that it is relatively easy to influence a client's responses. For example, saying the word *good* after each response can increase the overall number of responses on the Rorschach by as much as 50% (Hersen & Greaves, 1971). Similarly, examiners who were told that more experienced examiners elicited a greater proportion of human than animal responses actually produced this pattern from examinees, even though the examiners believed they were providing a standard administration (Exner, Leura, & George, 1976). These findings are consistent with the view that respondents are particularly responsive to subtle influences when attempting to create clarity in an ambiguous situation like Rorschach testing. However, if the fluctuations in administration style are minor, they are unlikely to influence a person's responses significantly. In general, examiners should minimize the variations in their administration procedures as much as possible. The next sequence of steps is derived from Exner (2003).

Step 1: Introducing the Respondent to the Technique

One of the most important goals an examiner must initially achieve is to allow the examinee to feel relatively comfortable with the testing procedure. Achieving this goal is complicated by the fact that tests in most cultures are associated with anxiety. Although in some cases, an increase in anxiety may provide some information that cannot be obtained when the subject is relaxed, anxiety is usually regarded as a hindrance. Typically, anxiety interferes with a person's perceptions and with the free flow of fantasy, both of which are essential for adequate Rorschach responses. Thus, subjects should be as relaxed as possible. Their relaxation can be enhanced by giving a clear introduction to the testing procedure, obtaining personal history, answering questions, and generally avoiding any behavior that might increase the clients' anxiety. In describing the test, examiners should emphasize relatively neutral words, such as *inkblot*, *interests*, or *imagination*, rather than potentially anxiety-provoking words, such as *intelligence* or *ambiguous*.

For the most part, any specific information regarding what clients should do or say is to be avoided. The test situation is designed to be ambiguous, and examiners should avoid any statements that might influence the responses. If respondents push for more detailed information about what they should do or what their responses may mean, they should be told that additional questions can be answered after the test is completed.

Step 2: Giving the Testing Instructions

Although some Rorschach systematizers recommend that the respondent tell the examiner "everything you see" (S. J. Beck, 1961), the Comprehensive System attempts to keep the task as ambiguous as possible. Thus, Exner (2003) instructed that the examiner hand the respondent the first card and ask, "What might this be?"

Commentary on or discussion of the cards by the examiner should be avoided as much as possible. At times, it might be acceptable to briefly describe how the designs

were made, or, if questioned regarding what one is supposed to see, the examiner might state, "People see all sorts of things in the blots." Comments from the examiner that indicate the quantity or type of response, or whether the subject can turn the cards, should be strictly avoided. If the client asks specific questions, such as the type of responses he or she is supposed to give or whether he or she can turn the cards, the examiner might reply that it is up to him or her to decide.

The main objective is to give the respondent maximum freedom to respond to the stimuli in his or her own manner. To enhance this, Exner (2003) strongly recommended that the subject and the examiner not be seated face to face but rather side by side, to decrease the possible influence of the examiner's nonverbal behavior. The overall instructions and testing situation should be designed both to keep the task as ambiguous as possible and to keep examiner influence to a minimum. Note that the examinee should be encouraged actually to hold the cards.

Step 3: The Response (Association) Phase

Throughout the testing procedure, the basic conditions of step 2 should be adhered to as closely as possible. However, specific situations often arise as examinees are free-associating to the Rorschach designs. If a respondent requests specifics on how to respond or asks the examiner for encouragement or approval, examiners should consistently reply that the subject can respond however he or she likes. Sometimes the idea that there are no right or wrong answers might be mentioned.

The examiner should time the interval that begins when the respondent first sees the card and ends when he or she makes an initial response as well as the total time the respondent spends with each card. These measurements can be helpful in revealing the general approach to the card and the possible difficulties in coming up with responses. Cards II, III, and V are generally considered relatively easy to respond to and, as a result, usually have shorter reaction times. In contrast, cards VI, IX, and X typically produce the longest reaction times. Because overt timing of subjects' responses is likely to produce anxiety, any recording should be done as inconspicuously as possible. It is recommended that, rather than using a stopwatch, the examiner glance at a watch or clock and record the minute and second positions for the initial presentation, the first response, and the point at which the subject hands the card back to the examiner.

The average number of responses is 22.32 (average range = 17–27). Validity can be compromised with a low number of responses (under 14) and may be questionable with a high number of responses (more than 42). Exner (2003) built in some safeguards to protect against unusually short or extremely long protocols. A client who produces an extremely brief protocol (fewer than 14 responses) should be retested immediately and provided with a clearer request to provide more responses (Exner, 2003). If a client provides more than five responses to the first inkblot, the examiner should remove the inkblot. On all subsequent inkblots, the same procedure should be used whenever the client provides five or more responses. However, if fewer than five responses to the first inkblot are given, no other limits on either the first inkblot or any later inkblots should be provided.

Exner (2003) stressed that all responses must be recorded verbatim. To simplify this process, most clinicians develop a series of abbreviations. A set of abbreviations used

throughout all the Rorschach systems consists of the symbols (\vee , $<$, \wedge , $>$) in which the peak indicates the angle of the card. It is also important to note any odd or unusual responses to the cards, such as an apparent increase in anxiety, wandering of attention, or acting-out on any of the percepts.

Step 4: Inquiry

The inquiry should begin after all 10 cards have been administered. Its purpose is to collect the additional information required for an accurate coding of the responses. It is intended to clarify the responses that have already been given, not to obtain new responses. The information needed from the inquiry phase should ensure that the examiner knows the location, content, and determinants for each response. (The other codes can be coded regardless, but information on these three codes needs to be explicitly gathered.) The inquiry should not end until this goal has been accomplished. Exner (2003) recommended that the instructions for the inquiry closely approximate these ones:

Now we are going to go back through the cards again. It won't take very long. I want to see the things that you said you saw and make sure that I see them like you do. We'll do them one at a time. I'll read what you said and then I want you to show me where it is in the blot and then tell me what there is there that makes it look like that to you, so that I can see it too, just like you did. Is that clear? (p. 59)

Following closely the general theme of the overall administration, the inquiry should not influence the examinee's responses. Thus, any questions should be as nondirective as possible. The examiner should begin by merely repeating what the respondent has said and then waiting. Usually the respondent begins to clarify his or her response. If this information is insufficient to clarify how to code the response (location, content, determinants), the examiner might become slightly more directive by asking "What about it made it look like [a percept]?" The examiner should *not* ask, "Is it mainly the shape?" or "How important was the color?" These questions are far too directive and are worded in a way that can exert influence on the respondent's descriptions of his or her responses. The examiner should consistently avoid leading the client or indicating how he or she should respond. Particular skill is required when clarifying a determinant that has been unclearly articulated but merely implied.

The outcome of a well-conducted inquiry is the collection of information sufficient to decide on coding for location, content, and determinants. If, on the location, information based on the client's verbal response is insufficient, the examiner should have the client point to the percept. An additional feature of the inquiry is to test the client's awareness of his or her responses. For example, does a strange percept represent coherent creativity, or does it reflect a lack of contact with the environment, with the individual perhaps having no awareness of the strangeness of his or her responses? The overall approach of the inquiry is to word questions in such a way as to be flexible without being too directive.

COMPREHENSIVE SYSTEM: CODING

Following administration, the next step is to code the different categories for each response. There is general agreement throughout the different Rorschach systems that these categories include at a minimum location, determinants, content, and popularity. The Comprehensive System also includes 15 Special Scores for responses, such as unusual verbalizations and aggressive movement. After these have been coded and tallied, a series of quantitative summaries, including six Special Indices, is created based on reorganizations of, and comparisons among, the scores on the different categories.

The subsections that follow merely list, outline, and define the coding categories. To achieve accurate coding (and scoring), it would be necessary to consult Exner's criteria (2003) or to use his workbook (Exner, 2001, *A Rorschach Workbook for the Comprehensive System*, 5th ed.), which includes specific coding criteria, tables, charts, and diagrams. The inclusion of specific scoring criteria is beyond the scope of this chapter. The focus here is on providing a key to interpretation that is concise, accountable, and clearly organized. The definitions and the accompanying tables serve to outline and briefly define the primary Comprehensive System factors.

Location and Developmental Quality

The *location* of the responses refers to the area of the inkblot that is used (Table 11.1). The location can vary from the use of the entire blot (whole response) to the use of an unusual detail (Dd). Unusual details are defined as location responses made by less than 5% of normative sample subjects. Exner (2003) also specified coding for Developmental Quality, which is determined by evaluating each location code in relation to its degree of integration. Table 11.2 presents the criteria used for coding the respective Developmental Quality codes. Thus, each location response is given both a designation for the specific area of the blot and a symbol to indicate the quality of that response.

Table 11.1 Symbols Used for Coding the Location of Rorschach Responses

Symbol	Definition	Criterion
W	Whole response	Where the entire blot is used in the response. All portions must be used.
D	Common detail response	A frequently identified area of the blot.
Dd	Unusual detail response	An infrequently identified area of the blot.
S	Space response	A white-space area is used in the response (scored only with another location symbol, as in WS, DS, DdS).

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Table 11.2 Symbols and Criteria Used for Developmental Quality

Symbol	Definition	Criterion
+	Synthesized response	Two or more objects are described as separate but related. <i>At least one</i> of the objects involved must have a specific form demand, or be described in a manner that creates a specific form demand (e.g., a dog walking among some bushes, a man with a funny hat on, an airplane flying through some clouds, the head of a little girl, she has a hair ribbon).
o	Ordinary response	An area of the blot is identified as a single object that has features that create a natural form demand or the <i>description of the object is such as to create</i> a specific form demand (e.g., a fir tree, a cat, a totem pole, a maple leaf, a bat, a flag, a man's head).
v/+	Synthesized response	Two or more objects are described as separate but related. <i>None of the objects</i> involved have a specific form demand and the articulation does not introduce a form demand for any of the objects (e.g., clouds coming together, some sort of bay with the vegetation around the shore, a rock and some dirt around it).
v	Vague response	An object is reported that has no specific form demand, <i>and the articulation does not introduce</i> a specific form demand for the object (e.g., a cloud, the sky, the colors of sunset, some ice).

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Determinants, Form Quality, and Organizational Activity

The term *determinants* refers to the style or characteristic of the blot to which the examinee responds, such as its shape, color, or texture (Table 11.3). Determinants are the reason an examinee gives that the blot looks like whatever he or she responded as having seen. The determinants also receive a coding for their level of form quality (Table 11.4). The form quality coding refers to how accurately the percept relates to the form of the inkblot. For example, an angel on Card I is considered to be an “ordinary” form quality response, which is empirically reflected in the fact that nonpsychiatric populations perceive it far more frequently than psychiatric patients do. Initially, examiners should give a percept its appropriate classification regarding its determinants. Then examiners should code the determinant for its relative form quality. Descriptions of the different form qualities are included in Table 11.4; however, for specific empirically derived form quality codings, examiners need to consult Exner's (2003) tables.

One relevant coding that should be added to all movement responses is the extent to which the movement is active versus passive. Active movement would include movements such as “fleeing” or “lifting,” whereas more passive movements might include “meditating” or “anchored.” Whether a movement is active or passive is designated

Table 11.3 Symbols and Criteria for Determinant Coding

Category	Symbol	Criterion
Form	F	<i>Form answers.</i> Used for responses based exclusively on the form features of the blot.
Movement	M	<i>Human movement response.</i> Used for responses involving the kinesthetic activity of a human or an animal or fictional character in human-like activity.
	FM	<i>Animal movement response.</i> Used for responses involving a kinesthetic activity of an animal. The movement perceived must be congruent to the species identified in the content. Animals reported in movement not common to their species should be coded as M.
	m	<i>Inanimate movement response.</i> Used for responses involving the movement of inanimate, inorganic, or insensate objects.
Chromatic color	C	<i>Pure color response.</i> Used for answers based exclusively on the chromatic color features of the blot. No form is involved.
	CF	<i>Color-form response.</i> Used for answers that are formulated primarily because of the chromatic color features of the blot. Form features are used but are of secondary importance.
	FC	<i>Form-color response.</i> Used for answers that are created mainly because of form features. Chromatic color is used but is of secondary importance.
	Cn	<i>Color naming response.</i> Used when the colors of the blot are identified by name and with the intention of giving a response.
Achromatic color	C'	<i>Pure achromatic color response.</i> Used when the response is based exclusively on the grey, black, or white features of the blot, when they are clearly used as color. No form is involved.
	C'F	<i>Achromatic color-form response.</i> Used for responses that are created mainly because of the black, white, or grey features, clearly used as color. Form features are used but are of secondary importance.
	FC'	<i>Form-achromatic color response.</i> Used for answers that are based mainly on the form features. The achromatic features, clearly used as color, are also included but are of secondary importance.
Shading-texture	T	<i>Pure texture response.</i> Used for answers in which the shading components of the blot are translated to represent a tactual phenomenon, with no consideration to the form features.
	TF	<i>Texture-form response.</i> Used for responses in which the shading features of the blot are interpreted as tactual, and form is used secondarily, for purposes of elaboration and/or clarification.
	FT	<i>Form-texture response.</i> Used for responses that are based mainly on the form features. Shading features of the blot are translated as tactual but are of secondary importance.

Continued

Table 11.3 Continued

Category	Symbol	Criterion
Shading-dimension	V	<i>Pure vista response.</i> Used for answers in which the shading features are interpreted as depth or dimensionality. No form is involved.
	VF	<i>Vista-form response.</i> Used for responses in which the shading features are interpreted as depth or dimensionality. Form features are included but are of secondary importance.
	FV	<i>Form-vista response.</i> Used for answers that are based mainly on the form features of the blot. Shading features are also interpreted to note depth and/or dimensionality but are of secondary importance to the formulation of the answer.
Shading-diffuse	Y	<i>Pure shading response.</i> Used for responses that are based exclusively on the light-dark features of the blot that are completely formless and do not involve reference to either texture or dimension.
	YF	<i>Shading form response.</i> Used for responses based primarily on the light-dark features of the blot, not involving texture or dimension. Form features are included but are of secondary importance.
	FY	<i>Form-shading response.</i> Used for responses that are based mainly on the form features of the blot. The light-dark features of the figure, not used to articulate texture or dimension, are included as elaboration and/or clarification and are secondary to the use of form.
Form dimension	FD	<i>Form-based dimensional response.</i> Used for answers in which the impression of depth, distance, or dimensionality is created by using the elements of size and/or shape of contours. No use of shading is involved in creating this impression.
Pairs and reflections	(2)	<i>The pair response.</i> Used for answers in which two identical objects are reported, based on the symmetry of the blot. The objects must be equivalent in all respects, but must not be identified as being reflected or as mirror images.
	rF	<i>Reflection-form response.</i> Used for answers in which the blot or blot area is reported as a reflection or mirror image because of the symmetry of the blot. The object or content reported has no specific form requirement, as in clouds, landscape, shadows, and so on.
	Fr	<i>Form-reflection response.</i> Used for answers in which the blot or blot area is identified as reflected or a mirror image, based on the symmetry of the blot. The substance of the response is based on form features, and the object reported as a specific form demand.

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Table 11.4 Symbols and Criteria for Coding Form Quality

Symbol	Definition	Criterion
+	Ordinary-elaborated	The unusually detailed articulation of <i>form</i> in responses that otherwise would be scored ordinary. It is done in a manner that tends to enrich the quality of the response without sacrificing the appropriateness of the form use. The answer is not necessarily original or creative; but, rather, it stands out by the manner in which form details are used and specified.
o	Ordinary	The common response in which general form features are easily articulated to identify an object. These are easy-to-see answers that have been reported by at least 2% of persons in the Form Quality data pool for <i>W</i> and <i>D</i> areas, or by at least 50 persons in the pool who responded to <i>Dd</i> areas. There is no unusual enrichment of the answer by elaboration of the form features.
u	Unusual	A low-frequency response in which the basic contours involved are appropriate for the response. These are uncommon answers that are seen quickly and easily by the observer.
2	Minus	The distorted, arbitrary, unrealistic use of form in creating a response. The answer is imposed on the blot structure with total or near total disregard for the contours of the area used. Often, substantial arbitrary lines or contours will be created where none exist.

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

with either an *a* (for active) or a *p* (for passive) superscript. The *a* and *p* designations are later scored and used for interpretation in the quantitative summaries (see the “Ideation Section” topic in the section titled “Structural Summary”).

In approximately 20% of all responses, more than one determinant is used to make a single response. These are referred to as *blends* and are designated by indicating the two (or more) determinants and placing a full stop (.) between them. The most important determinant is placed in front of the other determinant(s) and is considered the primary determinant. Less important determinants are placed after the primary one and are referred to as secondary or tertiary (if a third one is present).

A further code related exclusively to form determinants is the degree of Organizational Activity (Z) involved in creating the response. However, Organizational Activity is given only if at least one of these three criteria is present:

1. A *W* response, but only if its DQ coding is +, o, or v/+ (DQv responses are not coded for Organizational Activity).

2. Two or more separate objects are identified as being in some meaningful relationship.
3. White space is used in the response, integrated with other areas of the blot (responses that use *only* white space are not coded for Organizational Activity).

Specific converted weightings (ranging between 1 and 6) are given to organizational activity efforts for different types of responses and are provided in Exner (2003; see Table 8.4, p. 132). For example, the degree of organization required to integrate a whole response to Card I is considered to be much less (Z would equal only 1.0) than that required to integrate the much more fragmented details of Card X (Z would equal a much greater 6.5).

Content

The coding of content is based on the type and quantity of specific subjects that examinees perceive in their responses. Each Rorschach system uses different lists of content categories, although they all agree on basic contents such as human, human detail, and animal. Table 11.5 provides a listing of the content categories included in the Comprehensive System, with the symbol and criterion for each category.

When two or more content categories occur in the same response, they should both be coded and a comma should be placed between the two (or more) codings. If contents occur that are not on the list, they should be designated as idiographic (Id), and the unique name of the content should be written out.

Popular Responses

Comprehensive System Popular (P) coding refers to the presence of frequently perceived responses. Although different systems have somewhat varying lists of Populars, Exner (2003) used, as the cutoff for inclusion as a Popular, an occurrence of at least once in every three protocols from nonpsychiatric populations. Exner's list of Popular responses is detailed in Table 11.6.

Special Scores

The Comprehensive System also includes 15 Special Score categories that were developed to take into account unusual characteristics of the response, such as unusual verbalizations or inappropriate logic. These categories, along with their definitions, are listed in Table 11.7. The first four Special Scores (DV, DR, INCOM, and FABCOM) are coded at a Level 1 if the response is only mildly atypical and at a Level 2 if the response is more bizarre or unusual, indicating more likely cognitive slippage.

COMPREHENSIVE SYSTEM: SCORING THE STRUCTURAL SUMMARY

After the examinee's responses have been coded according to locations (and developmental quality), determinants (and form quality and organizational activity), contents

Table 11.5 Symbols and Criteria Used for Coding Content

Category	Symbol	Criterion
Whole human	<i>H</i>	For responses involving a whole human form. If the response involves a <i>real</i> historical figure, such as Napoleon, Joan of Arc, and so on, the content code <i>AY</i> should be added as a secondary code.
Whole human, fictional or mythological	<i>(H)</i>	For responses involving a whole human form that is fictional or mythological, such as clowns, fairies, giants, witches, fairy-tale characters, angels, dwarfs, devils, ghosts, science-fiction creatures that are humanoid, human-like monsters, silhouettes of human figures.
Human detail	<i>Hd</i>	For responses involving an incomplete human form, such as an arm, head, leg, fingers, feet, the lower part of a person, a person without a head.
Human detail, fictional or mythological	<i>(Hd)</i>	For responses involving an incomplete human form that is fictional or mythological, such as the head of the devil, the arm of a witch, the eyes of an angel, parts of humanoid science-fiction creatures, jack-o-lantern, and masks <i>except</i> animal masks.
Human experience	<i>Hx</i>	Usually coded as a secondary content for answers that clearly involve the attribution of a human emotion or sensory experience to the object(s) in the response, such as <i>two people who are in love looking at each other</i> , <i>a cat that is very sad</i> , <i>people who are angry at each other</i> , <i>a woman smelling something nasty</i> , <i>a very happy person</i> , <i>a man who is very excited</i> , <i>a person in great pain</i> . The attribution of the motion or sensory experience must be clear and unequivocal. Answers such as <i>people at a party</i> , <i>an angry-looking face</i> , <i>a mean-looking person</i> , <i>two people who look tired</i> are not coded <i>Hx</i> as the attribution is equivocal. <i>Hx</i> is scored as a primary content for formless <i>M</i> responses that involve the emotion or sensory experience such as love, hate, depression, happiness, sound, smell, fear, and so on. These answers will also include the use of <i>AB</i> as a special score.
Whole animal	<i>A</i>	For responses involving a whole animal form.
Whole animal, fictional or mythological	<i>(A)</i>	For responses involving a whole animal that is fictional or mythological, such as a unicorn, dragon, magic frog, flying horse, Black Beauty, Jonathan Livingston Seagull.
Animal detail	<i>Ad</i>	For responses involving an incomplete animal form, such as the hoof of a horse, claw of a lobster, head of a dog, animal skin.
Animal detail, fictional or mythological	<i>(Ad)</i>	For responses involving an incomplete animal form that is fictional or mythological, such as the wing of Pegasus, the head of Peter Rabbit, the legs of Pooh Bear, and all animal masks.
Anatomy	<i>An</i>	For responses in which the content is skeletal, muscular, or of internal anatomy such as bone structure, skull, rib cage, heart, lungs, stomach, liver, muscle fiber, vertebrae, brain. If the response involves a tissue slide, the content <i>Art</i> should be added as secondary.

Continued

Table 11.5 Continued

Category	Symbol	Criterion
Art	<i>Art</i>	For responses of paintings, drawings, or illustrations, either abstract or definite, art objects such as statues, jewelry, chandelier, candelabra, crests, badges, seals, and decorations. A feather seen worn as a decoration, often seen on Card VII, also should be coded as Art. In many responses coded for Art a second content will also be coded, such as a painting of two dogs would be <i>Art</i> , <i>A</i> , a sculpture of two witches would be <i>Art</i> , (<i>H</i>), a caricature of two people bending over would be <i>Art</i> , <i>H</i> .
Anthropology	<i>Ay</i>	For responses that have a specific cultural or historical connotation such as totem, roman helmet, Magna Carta, Santa Maria, Napoleon's hat, Cleopatra's crown, arrowhead, prehistoric axe, an Indian war bonnet.
Blood	<i>Bl</i>	For responses of blood, either human or animal.
Botany	<i>Bt</i>	For responses involving any plant life such as bushes, flowers, seaweed, trees or parts of plant life, such as leaves, petals, tree trunk, root, bird's nest.
Clothing	<i>Cg</i>	For responses involving any article of clothing such as hat, boots, belt, dress, necktie, jacket, trousers, scarf.
Clouds	<i>Cl</i>	For responses used specifically for the content cloud. Variations of this category, such as fog or mist, are coded <i>Na</i> .
Explosion	<i>Ex</i>	For responses involving a blast or explosion, including fireworks.
Fire	<i>Fi</i>	For responses of fire or smoke.
Food	<i>Fd</i>	For responses used for any edible common for humans, such as fried chicken, ice cream, fried shrimp, vegetables, cotton candy, chewing gum, steak, a filet of fish, or for animals eating a food that is natural for their species, such as a bird eating a worm or insect.
Geography	<i>Ge</i>	For responses used for the response of a map, specified or unspecified.
Household	<i>Hh</i>	For responses used that include household items, such as bed, carving knife, chair, cooking utensils, cup, garden hose, glass, lamp, lawn chair, plate, rug (<i>animal skin rug should be coded Ad and Hh entered as a secondary content</i>), silverware. Some items coded <i>Hh</i> will also be coded as <i>Art</i> , such as candelabra, chandelier, or artistic pieces such as a centerpiece bowl.
Landscape	<i>Ls</i>	For responses that involve landscape, such as mountain, mountain range, hill, island, cave, rocks, desert, swamp, or seascapes, such as coral reef or underwater scene.
Nature	<i>Na</i>	For responses used for a broad variety of contents from the natural environment that are not coded as <i>Bt</i> or <i>Ls</i> , such as sun, moon, planet, sky, water, ocean, lake, river, ice, snow, rain, fog, mist, rainbow, storm, tornado, night, raindrop.

Table 11.5 Continued

Category	Symbol	Criterion
Science	Sc	For responses that are associated with, or are the direct or indirect products of science or science fiction, such as airplanes, buildings, bridges, cars, light bulb, microphone, motorcycles, motors, musical instrument, radar station, road, rocket ships, ships, space ships, trains, telescope, TV aerial, weapons, and so on.
Sex	Sx	For responses involving sex organs or activity of a sexual nature, such as penis, vagina, buttocks, breasts (except when used to identify the sex of a human figure), testes, menstruation, abortion, intercourse. <i>Sx</i> is usually scored as a secondary content. Primary contents are typically <i>H</i> , <i>Hd</i> , or <i>An</i> .
X-ray	Xy	For responses used specifically for the content of x-ray and may include either skeleton or organs. When <i>Xy</i> is coded, <i>An</i> is <i>not</i> included as a secondary code.

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

(and populars), and Special Scores, they are listed and rearranged into frequency summaries and quantitative formulas. The quantitative formulas comprise various ratios, percentages, and derivations. These formulas reflect the proportions of, and comparisons among, various Rorschach factors. After the quantitative formulas have been calculated, they become the primary focus on which Rorschach interpretations are made. Exner (2003) categorized the formulas into a Core section, followed by sections for Ideation, Affect, Mediation, Processing, Interpersonal, Self-Perception, and Special Indices (Depression Index, Obsessive Style Index, etc.). These sections provide a convenient way to thematically organize the different interpretations. The calculations and descriptions are presented by Exner. The various scorings, frequencies, and formulas can be conveniently summarized on a commercially available record form that includes a *Structural Summary Blank* as well as a *Constellation Worksheet* for calculating the Special Indices.

COMPREHENSIVE SYSTEM: INTERPRETATION

The following description of interpretive information is meant to serve as a reference guide to alert Comprehensive System interpreters to a potentially wide range of possible interpretive hypotheses. Although the format is as concise as possible, interpreters should be aware of the tremendous variety inherent in most Rorschach data. Effective interpreters should also have this variety reflected in the wide number of possible interpretive hypotheses they generate. A mere labeling or simplistic “sign” approach should be avoided. Rather, clinicians must begin and end by continually being aware of the total overall configuration of the data. For example, the same number of C responses in

Table 11.6 Popular Responses Used in the Comprehensive System Plus the Proportions of Each Appearing in Samples of Nonpatient and Patient Protocols

Card	Location	Criterion	Nonpatient %	Patients %
I	<i>W</i>	Bat, with the true apex of the blot being identified as the top portion of the bat, and always involving the whole blot.	48	38
I	<i>W</i>	Butterfly, with the true apex of the blot being identified as the top portion of the butterfly, and always involving the whole blot.	40	36
II	<i>D1</i>	Animal, specifically identified as bear, dog, elephant, or lamb. The response is usually the head or upper body, but responses involving the whole animal are also coded <i>P</i> .	34	35
III	<i>D9</i>	Human figures or representations thereof such as dolls, caricatures, and so on. If <i>D1</i> is used as two human figures, <i>D7</i> or <i>Dd31</i> should not be reported as part of the human figure.	89	70
IV	<i>W</i> or <i>D7</i>	Human or human-like figure such as giant, monster, science-fiction creature, and so on. Animal figures are not coded as <i>P</i> .	53	41
V	<i>W</i>	Butterfly, with the true apex of the blot being identified as the top portion of the butterfly, and always involving the whole blot. The whole blot <i>must</i> be used.	46	43
V	<i>W</i>	Bat, with the true apex of the blot being identified as the top portion of the bat, and always involving the whole blot.	36	38
VI	<i>W</i> or <i>D1</i>	Animal skin, hide, rug, or pelt. Often, the skin, hide, or pelt will be included in the description of a whole animal, such as a cat or fox, in natural or unnatural form. The decision about whether to code <i>P</i> in these responses is based on whether the skin or hide is actually mentioned or clearly implied.	87	35
VII	<i>D1</i> or <i>D9</i>	Human head or face, specifically identified as female, child, Indian, or with gender not identified. This Popular is usually embedded in answers given to the larger areas, <i>D1</i> , <i>D2</i> , or <i>Dd23</i> . If <i>D1</i> is used, the upper segment (<i>D5</i>) is typically identified as hair or a feather. If the response includes the entire <i>D2</i> or <i>Dd23</i> areas, <i>P</i> is coded only if the head or face is restricted to the <i>D9</i> area.	59	47
VIII	<i>D1</i>	Whole animal figure, usually of the canine, feline, or rodent varieties, with the head of the animal adjacent to the <i>D4</i> area.	94	91

Table 11.6 Continued

Card	Location	Criterion	Nonpatient %	Patients %
IX	D3	Human or human-like figures such as witches, giants, science-fiction creatures, monsters, and so on.	54	24
X	D1	Spider with all appendages restricted to the D1 area.	42	34
X	D1	Crab with all appendages restricted to the D1 area. Other variations of multilegged animals are not P.	37	38

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

two protocols can easily have quite different meanings, depending on the implications from, and interactions with, other aspects of the Rorschach data.

The typical sequence for Rorschach interpretation should follow the general conceptual model for testing developed by Wright (2010) and discussed in Chapter 1 (see Figure 1.1). The model requires that clinicians initially take a tentative stance toward the data (phase 5). The purpose of this stage is to develop as many tentative hypotheses as possible, based on the quantitative data, verbalizations, and client history. The number and accuracy of these hypotheses depend on the individual richness of the data as well as on the individual skill of the clinician. The next stage (phase 6) relates to evaluating these hypotheses and rejecting, modifying, or confirming them as necessary, followed by the integration of the hypotheses into a meaningful and accurate description of the person (phase 7). When this has been accomplished, clinicians can integrate the Rorschach interpretations into the overall report itself.

In the description of different interpretive hypotheses, continual reference is made to “high” and “low” scores. These relative weightings are based on extensive adult normative data that have been accumulated on the Comprehensive System. Note that normative scores (means and standard deviations), derived from Exner and Erdberg (2005), have been included in each of the interpretive sections for quick reference. The relative validity of various categories is indicated and is based on Mihura et al. (2013). It should be noted that the validity ratings were not done for the individual codes but were provided exclusively for the ratios, percentages, and derivations.

Clinicians interested in child and adolescent assessment (ages 6–16) can consult the much more extensive age-based norms for children found in Exner (2003) and Exner and Weiner (1995). Clinicians interested in cross-national interpretations (and many argue general, nonpsychiatric populations) should refer to the international normative reference group published by Meyer et al. (2007).

The sequence of presenting interpretive information is first organized around specific coding categories (Location, Determinants, Contents, Special Scores). These are followed by the sections in the Structural Summary. The calculations in the Structural Summary begin with the Core section and then proceed to the sections for Ideation, Affect, Mediation, Processing, Interpersonal, Self-Perception, and Special

Table 11.7 Symbols and Descriptions for Special Scores

Special Score (Symbol)	Description
Deviant Verbalization (DV)*	Verbalizations associated with a response, which are odd and suggest some form of cognitive slippage has occurred, such as through neologisms or redundancies (e.g., “pair of two”).
Deviant Response (DR)*	Responses that involve a longer segment of the response than verbalizations, such as through inappropriate phrases or circumstantial responses that are long, rambling, and unrelated to the inkblot.
Incongruous Combination (INCOM)*	Images that have been inappropriately merged into a single object.
Fabulized Combination (FABCOM)*	Implausible relationships between two or more objects identified in the inkblot.
Contamination (CONTAM)	Two or more impressions that have been inappropriately fused together.
Inappropriate Logic (ALOG)	Spontaneously offered justification of the response using strained logic.
Perseveration (PSV)	Providing either an identical or almost identical response two or more times in a row, or seeing the same object repeatedly (“There’s that man again”).
Abstract Content (AB)	Symbolic representation is given to the content.
Aggressive Movement (AG)	Any movement response that is clearly aggressive.
Cooperative Movement (COP)	Any movement response that is clearly cooperative.
Morbid Content (MOR)	Content is characterized by death or damage, or is designated as being dysphoric.
Good Human Representation (GHR)	Positive representation of humans (i.e., Pure Human coding with +, o, or u Form Quality; see Exner, 2003, Table 9.1).
Special Score (Symbol)	Description
Poor Human Representation	Poor representations of humans (e.g., human responses that are Form Quality –; see Exner, 2003, Table 9.1, p. 144).
Personal (PER)	Reference to personal knowledge or experience is used to justify or clarify a response.
Color Projection (CP)	Identification of an achromatic portion of an inkblot as being colored.

*These Special Scores are rated as either Level 1, indicating a mild to modest level of cognitive slippage, or Level 2, indicating that level of cognitive slippage is moderate to severe.

Source: From *The Rorschach: A Comprehensive System, Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, Hoboken, NJ: Wiley. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Indices. These later groupings should provide a conceptually consistent means of organizing relevant interpretive material around functional domains, thereby enabling the different interpretations to be more easily integrated into the psychological report. For example, if a practitioner is interested in understanding issues related to interpersonal relationships, he or she can note the Rorschach data relevant to this area of functioning. Similarly, information related to dealing with affect can be noted in the section on affect. These interpretations can then be compared, contrasted, and modified with other assessment material on these dimensions. Table 11.8 outlines the different interpretive categories in the order in which they are presented for interpretation. Again, the Comprehensive System organizes the data into these eight clusters (see Exner, 2003, Table 13.2, p. 225):

1. Controls and Stress Tolerance
2. Situational Stress
3. Affective Features
4. Self-Perception
5. Information Processing
6. Mediation
7. Ideation
8. Interpersonal Perception

It is believed that the structure provided in the section titled “Comprehensive System: Interpretation” should be relatively easy to follow and understand since it is organized according to the scoring sequence. The organization based on the scoring sequence should provide clinicians with a means of developing some working knowledge of the Rorschach. The listing of the various clusters of the Rorschach serves as a *de facto* Comprehensive System–based theory of personality in that clinicians can conceptualize cases based on these essential aspects of functioning.

The process of reading through the many interpretations in the remainder of the chapter can be tedious because of the sheer quantity. To deal with the quantity of interpretations, it is recommended that the practitioner initially skim over the different sections and interpretations. Next, he or she might obtain a Rorschach protocol by actually administering and coding/scoring a Rorschach, requesting one from a colleague, or using one from one of Exner’s books. The practitioner can then go through each of the different categories and generate hypotheses based on the client’s results. The hypotheses can be integrated into a description of the person, based on domains measured by the Rorschach variables. This sequence would make the information relevant and engaging as well as enhance the development of actual clinical skills.

Location

In general, the area of the inkblot to which examinees choose to respond is a reflection of the overall style in which they approach their world. This is especially true for the manner in which they confront uncertainties and ambiguities in their

Table 11.8 Scoring and Interpretative Domains for the Comprehensive System

Location	Ratios, Percentages, Derivations
Whole Response (W)	Core Section —frequency data (taken from previous sections includes total number of responses plus each of the frequencies of the determinants) and the following nine formulas:
Common Detail (D)	1. Lambda (L)
Unusual Detail (Dd)	2. Experience Balance or Erlebnistypus (EB)
Space (S)	3. Experience Actual (EA)
Developmental Quality (DQ)	4. Experience Pervasive (EBPer)
	5. Experience Base (eb)
	6. Experience Stimulation (ES)
	7. D Score (D)
	8. Adjusted es (Adj es)
	9. Adjusted D score (Adj/D)
Determinants	Ideation Section —frequency data for M–, M, number of Level 2 responses, WSum6, and M with no FQ. In addition:
Form (F)	1. Active: Passive Ratio (a: p)
Human Movement (M)	2. M Active: Passive Ratio (Ma: Mp)
Animal Movement (FM)	3. The Intellectualization Index [2AB + (Art + Ay)]
Inanimate Movement (m)	
Color Chromatic (C)	Affect Section —frequency of Pure C, S, and CP and the following three formulas:
Color Achromatic (C9)	1. Form-Color Ratio [(FC; CF + C)]
Shading—Texture (T)	2. Affective Ratio (Afr)
Shading—Dimension (V)	3. Complexity Index (Blend; R)
Form Dimension (FD)	Mediation Section —number of Popular responses, the total number of S- responses, and the following percentages:
Pairs (2) and Reflections (Fr)	1. Form Appropriate Extended (XA+%)
Organizational Activity (Z)	2. Form Appropriate-Common Areas (WDA%)
	3. Distorted Form (X2%)
	4. Conventional Form
	5. Unusual Form (Xu%)
Content	Processing Section —frequency data for Organization Activity (Zf), Perseverations (PSV), Developmental Quality+ (DQ+), Developmental Quality-v (DQv), and three ratios:
Human and Human Detail (H, Hd)	1. Economy Index (W; D; Dd)
Animal and Animal Detail (A, Ad)	2. Aspirational Ratio (W; M)
Anatomy and X Ray (An, Xy)	3. Processing Efficiency (Zd)
Food (Fd)	
Popular Responses	
Special Scores	
Deviant Verbalizations (DV)	
Deviant Responses (DR)	
Incongruous Combinations (INCOM)	
Fabulized Combination (FABCOM)	
Contamination (CONTAM)	
Inappropriate Logic (ALOG)	
Perseveration (PSV)	
Abstract Content (AB)	
Aggressive Movement (AG)	
Cooperative Movement (COP)	
Morbid (MOR)	
Good Human Representation (GHR)	
Poor Human Representation (PHR)	
Personal (PER)	
Color Projection (CP)	

Table 11.8 Continued

Interpersonal Section —frequencies of Cooperative Movements (COP), Aggressive Movements (AG), Food Contents, sum of Pure H, number of Perseverations (PER), ratio of Good Human to Poor Human Representation (GHR:PHR), Sum T, and active:passive (a:p), and the following two formulas: 1. Interpersonal Interest $(H) + (H) + Hd + (Hd)$ 2. Isolation Index $Bt + 2Cl + Ge + Ge + Ls + 2NA/R$	response, sum Form Dimension responses, sum morbid content, sum Anatomy, sum X ray, sum V, ratio of Pure H; $(H) + Hd + (Hd)$, and the: 1. Egocentricity Index $[3r + (2)/R]s$
Self-Perception Section —Sum Form-reflection and reflection-Form	Special Indices Perceptual Thinking Index (PTI) Depression Index (DEPI) Coping Deficit Index (CDI) Suicide Constellation Index (S-CON) Hypervigilance Index (HVI) Obsessive Style Index (OBS)

lives. For example, one person might perceive only the most obvious and concrete aspects of a situation, whereas another might avoid important aspects of a stimulus by focusing on small details and neglecting potentially more significant issues. An analysis of Rorschach locations does not provide information regarding why people approach their world in a certain manner; rather, it is limited to a description of their particular style.

Rorschach locations can be divided into usual and unusual features, depending on the area of the inkblot that is used. Frequently used locations, if they are within the normal number and of good quality, usually reflect good ties with reality, intelligence, ambition, good reasoning, and an ability to generalize. Unusual locations involving rarely used areas of the blot are associated with neurotic symptomatology, such as fears, anxiety, and obsessive or compulsive tendencies. An extreme use of unusual features may reflect more serious psychopathology (Exner, 2003). It should be noted that while hypotheses are presented below for different uses of location, data supporting these interpretive hypotheses are not currently available, so they should be made with extreme caution.

Whole Response (W)

The whole response is related to the degree to which a person can interact in an efficient and active manner with his or her environment. This is particularly true if the quality and organization of the responses are good. Whereas whole responses occur with the greatest frequency in children from 3 to 4 years of age (Exner & Weiner, 1995), there is a gradual decline in later childhood and adolescence until 30% to 40% of normal adult responses are wholes. The average adult ratio of whole:detail is approximately 1:2 (refer also to interpretation of W:M and W:D:Dd formulas).

High W ($M = 9.10$, $SD = 3.70$)

- High intellectual activity, good synthesizing ability, good abstract reasoning.
- Good ties with reality.

- Good problem solving abilities.
- The first two interpretations are dependent on the W responses using a high level of organizing activity (check proportion of W+ responses); the more the person can organize the response, the stronger the above interpretations.
- W occurs with greatest frequency for Cards V, I, IV, and VI and with lowest frequency for Cards X, IX, III, and VIII; W responses for the latter cards require significantly greater organizational activity.

Low W

- Possible depression or anxiety.
- If the frequency, quality, and complexity are low, more serious levels of maladjustment are indicated, such as intellectual deterioration possibly related to brain damage or mental retardation.

Common Detail (D)

Rorschach (1921/1941) originally conceptualized the D response as reflecting the degree to which a person perceives and reacts to the obvious aspects of a situation. This conceptualization is supported by more recent normative data in which adult nonpsychiatric groups and outpatients gave 62% and 67% of their responses, respectively, as D, whereas inpatient patients without schizophrenia and inpatient patients with schizophrenia gave 46% and 47% of their responses, respectively, as D (Exner, 1974). D tends to be most frequent for Card X (Exner, 2003). The proportion of D is lowest in young children and gradually increases with age (Ames, Metraux, Rodell, & Walker, 1974). Any interpretations relating to D should take into account the fact that a greater number of R is likely to increase the relative proportion of D when compared with other locations (also refer to the W:D:Dd ratio).

High D ($M = 12.66$, $SD = 4.75$)

- Overemphasis on concrete, obvious aspects of situations, sacrifices full use of intellectual potential by merely focusing on the safe and obvious rather than probing into the more novel and unusual.
- If D+ is high, an excellent level of functioning and a concern with precision are likely.
- If D is high but the quality of responses is low, a severe level of maladjustment is indicated.

Low D

- High experience of stress (if D is low along with a high Dd).
- Poor perceptual abilities possibly consistent with cerebral impairment.

Unusual Detail (Dd)

The Dd response is considered to represent a retreat from a person's environment by focusing on details rather than either perceiving the whole situation or noticing the

more obvious elements of the environment. A clinician would expect the number of Dd responses to comprise approximately 6% of the total R for a normal adult. However, Dd is frequently higher in the protocols of normal children and adolescents. For individuals with schizophrenia or severely impairing compulsivity, the proportion of Dd can increase to 25% or more (Exner, 1974). When Dd is in good proportion to W and D, a healthy adjustment, in which a person combines initiative with an appropriate ability to withdraw, is reflected.

High Dd ($M = 1.60$, $SD = 2.06$)

- Need to pull back from the ambiguities that may be contained in a whole response.
- Represents a person's attempt to narrow perceptions of the environment.
- Focus on the details of a situation in an attempt to reduce anxiety and exert more control over perceptions; consistent with a compulsive style.
- Rigid thought processes since the thought processes may not be flexible enough to take into account the ambiguities and complexities of whole responses.

Space (S)

A high number of S responses (three or more) is associated with negativism, difficulty in handling anger, and oppositional tendencies (Exner, 1993, 2003). If S responses are high (three or more) and occur with poor form quality and/or poor primitive movements, a clinician should consider the presence of anger, hostility, and potential acting out (Exner, 1993). There is support for interpretations based on this variable, but only when the space is used as the primary content of the response, as opposed to integrating space as a secondary characteristic of the blot (DeKoninck & Crabbe-Decleve, 1971; Dumitrascu, Mihura, & Meyer, 2011). That is, a response that uses white space in the middle of the blot to represent a spaceship would be interpretable in this way, whereas using white spaces to represent eyes on a face would not. It should be noted that interpreting the space content outside of this distinction has not been shown to have any validity for oppositionality or anger (Mihura et al., 2013) ($M = 2.37$, $SD = 1.97$).

DQ+ and DQv

Developmental quality scores relate to a person's relative ability to analyze and synthesize information. A high DQ+ (above 9 or 10) is consistent with more intelligent, complex, and sophisticated persons. However, this greater complexity does not necessarily mean that the person is well adjusted or even that his or her cognitions are accurate (see Zd for an index of both efficiency and accuracy). A number of disorders are characterized by quite complex cognitive operations, yet they are not well adjusted. In contrast to DQ+, a higher proportion (three or more) of low Developmental Quality (DQv) responses indicates persons who are immature and less sophisticated (e.g., children, those with neuropsychological impairment or intellectual disabilities; Exner, 1993, 2003). There is good support for both of these interpretations (Mihura et al., 2013) ($M = 8.43$, $SD = 3.07$, and $M = 0.37$, $SD = 0.72$).

Determinants

Because the majority of research has been done on the determinants, they are frequently seen as the core of the Rorschach data. An analysis of a person's determinant scores shows the psychological activity that he or she engaged in while the response was being created. This analysis examines his or her unique styles of perception and thinking and how these interact with one another. In general, research has isolated specific details of the determinants that could possibly lure the clinician into a rigid and potentially inaccurate "single sign" approach. Again, a Rorschach interpreter should focus on the interaction among a large number of variables to modify, confirm, or reject tentative hypotheses derived from any single determinant score.

Form (F)

The amount of pure F in a protocol has generally been used to indicate the extent to which the person can remove affect from a situation. The presence of form in a response represents a certain degree of respect for the standards of the environment and reflects intact reasoning abilities. It is seen both as related to attention and concentration and as an index of affective control or delay (Exner, 1993, 2003). This is reflected in the fact that inpatients with schizophrenia have a relatively higher percentage of Fu and F- responses than other groups. However, pure F is higher among those with paranoid schizophrenia than those with other types of schizophrenia (Rapaport, Gill, & Schafer, 1968), reflecting their greater degree of organization and caution. In addition, individuals with schizophrenia have increases in pure F following treatment (Exner, 1993), and a higher level of pure F for those with schizophrenia is associated with a better prognosis (Exner & Murillo, 1977). The presence of a pure F response does not necessarily mean that no conflict is present but rather that the person is able to suspend temporarily the affect associated with a conflict. Conversely, people in emotional turmoil are likely to produce a significantly lower number of pure form responses, reflecting their inability to remove their affect from their experience. (See also interpretations for Lambda and the percentages in the "Mediation section" of the chapter: X+%, F+%, X-%, S-%, Xu%.)

High Pure F ($M = 7.91$, $SD = 3.70$)

- Defensive, constricted.
- Good ability to deliberately suspend or control affect.
- Pure F increases for persons who have been given some prior knowledge of the purpose of the test or who are requested to respond as quickly as possible.

Low Pure F

- Level of turmoil is likely to be sufficiently high to prevent screening out affective response to a situation.
- Those with acute schizophrenia have difficulty reducing their level of affect and have a low number of pure F responses (Exner & Murillo, 1973). Also, organic disorders in which there is difficulty controlling impulses have a low number of pure F responses (Exner, 1974).

Human Movement (M)

Probably more research has been done on the M response than on any other Rorschach variable. Most of this research is consistent in viewing M as reflecting inner fantasies connected to the outside world. More specifically, M represents the bridging of inner resources with reality, or what might be described as “internalization of action” (Exner, 1993, 2003). It includes inner thinking, planning, imagination, and even empathy. M is also an inhibitor of outward behavior, even though that inhibition may be only temporary. A high proportion of M responses has been associated with creativity (Dudek, 1968) and introverted thinking (Kunce & Tamkin, 1981), and there is a close relationship between M and daydreaming (Dana, 1968; Page, 1957). Schulman (1953) showed M’s relation to abstract thinking in that a high number of M responses reflected both an active inner process and a delay in expressing behavior. Thus, M– can be generally understood as involving deliberate inner experience. In its positive sense, M can indicate good ego functioning, ability to plan, impulse control, and ability to withstand frustration. In a more negative vein, it can suggest an overdeveloped fantasy life. In general, there is excellent support for interpreting M in this way (Mihura et al., 2013).

While interpreting M, it is important to look carefully at the different components of the response. For example, does the movement involve conflict or cooperation? A high number of aggressive movements has been shown to reflect a person who is generally more aggressive and who typically perceives relationships as characterized by aggressiveness (Exner, 1983), and there is some support for this interpretation (Mihura et al., 2013). The degree of passivity in the movement is also likely to suggest that the person has more dependent and passive behaviors external to the test situation (Exner & Kazaoka, 1978), though it should be noted that there is not strong evidence to support this line of interpretation (Mihura et al., 2013). Specific interests might be projected into the movement responses, such as the increased number of dance movements perceived in the protocols of physical education and dance students (Kincel & Murray, 1984). The clinician should also consider other data both from within the test and external to it. Further elaboration regarding M, especially as it relates to the person’s degree of control of impulses, can be derived by referring to the EB and EA ratios.

High M ($M = 4.83$, $SD = 2.18$)

- High intelligence (especially if a high number of M are present).
- High creativity.
- Good abstract reasoning.
- Introverted thinking means of processing information.
- Capacity to delay impulses.
- High investment in fantasy life.
- With a high number M+ responses, good prognosis for psychotherapy.
- With high number of M– responses, distorted interpersonal perceptions (e.g., high M– responses have been found among patients with mania and those with psychotic perceptions of others). There is good support for this interpretation (Mihura et al., 2013).

Low M

- Possible depression, consistent with having a difficult time using inner resources.
- High impulsivity.
- Dementia.
- Rigidity, difficulty accepting and adjusting to change.
- Low empathy, lack of imagination.
- Poor prognosis for psychotherapy due to rigidity, low empathy, and poorly developed inner life.

Animal Movement (FM)

Whereas human movement responses serve to mediate between the inner and outer environment, animal movement has been considered to reflect more unrestrained emotional impulses over which there is less ego control. The impulses are more urgent, more conscious, and provoked by situations beyond the person's control. These observations are reflected in the higher number of FMs in children (Ames et al., 1974) and older adults (B. Klopfer, Ainsworth, Klopfer, & Holt, 1956). The number of FM responses for children (ages 8–16) typically ranges from 3.0 to 3.5, whereas adults have an average of 4.0 (Exner & Erdberg, 2005). However, although significant and in the expected direction, effect sizes were small for this interpretation, showing minimal evidence to support this category (Mihura et al., 2013).

High FM ($M = 4.04$, $SD = 1.90$)

- Thoughts and feelings are beyond the person's control.
- Unrestrained impulses, difficulty delaying gratification, rarely plan toward long-term goals.
- If high number of aggressive FM responses, possibly assaultive.
- Highly defensive with use of intellectualization, rationalization, regression, and substitution as primary means of reducing anxiety.

Low FM

- Overly inhibited in expressing emotions.
- May deny basic needs (associated with decreased energy level among children).

Inanimate Movement (m)

Similar to FM, the number of inanimate movement responses also provides an index of the extent to which persons are experiencing drives or life events that are beyond their ability to control, causing mental agitation. The drives reflected by m threaten people's adjustment in that they are helpless to effectively deal with them (B. Klopfer et al., 1956). This helplessness is usually related to interpersonal activities (Hertz, 1976; B. Klopfer et al., 1956; Z. Piotrowski, 1957, 1960). The average number of m responses for adult nonpatients was 1.6 ($SD = 1.34$; Exner & Erdberg, 2005). The view that m represents threat from the external world is supported by the observation that sailors at sea produced significantly more m during a severe storm (Shalit, 1965).

This is also consistent with the finding that normal subjects exposed to uncontrollable laboratory-induced stress (McCown, Fink, Galina, & Johnson, 1992) and those given amphetamines (W. Perry et al., 1995) had temporary increases in *m*. Similarly, paratroop trainees had an increase in *m* just before their first jump (Armbuster, Miller, & Exner, 1974), as did elective-surgery patients just before surgery (Exner, 1993; see also the interpretation of experience base [eb]). In contrast, Piotrowski and Schreiber (1952) found no *m* scores in the records of patients treated successfully. Thus, a number of studies support the interpretations for this scoring category (Hiller et al., 1999), and the effect size of this work is medium, which is strong evidence for these interpretations (Mihura et al., 2013). To gain a more complete understanding of the individual meaning of *m*, clinicians should investigate the possible resources and the characteristic means of resolving conflict by looking at *M*, sum *C*, frequency of *D* and *S*, and the accuracy of their perceptions as reflected in *F* + % and *X* + %.

High *m* ($M = 1.57$, $SD = 1.34$)

- Marked presence of conflict and tension.
- Perception that they are surrounded by threatening persons or events, feel unable to reconcile themselves with their environment.

Chromatic Color (C, CF, FC, Cn)

The manner in which color is handled reflects the style in which a subject deals with his or her emotions. If color dominates (*C*, *CF*, *Cn*), affect is likely to be poorly controlled and disorganized. In such cases, affect is disruptive, and the person could be expected to be emotional, labile, and overreactive. If the responses are more dominated by form (*FC*), affect will be more delayed, controlled, and organized. For example, it has been demonstrated that subjects who could effectively delay their responses in a problem-solving task had a higher number of *FC* responses in their protocols, whereas those who had difficulty delaying their responses had more *CF* and *C* responses (H. Gill, 1966; Pantle, Ebner, & Hynan, 1994). These researchers also found that a positive correlation exists between individuals having color-dominated responses and measurements of impulsiveness. However, if the number of color-dominated responses is used to determine impulsiveness, the implications of *D* scores, form quality, number of *Y* responses, and relative number of color-dominated responses ($FC:CF + C$) should also be taken into account. Furthermore, the chromatic cards produce a greater frequency of aggressive, passive, and undesirable contents than do the achromatic cards (Crumpton, 1956).

Adult nonpatients have between 1.5 to 2.5 times more form-dominated color than color-dominated responses [$FC/(CF + C)$]. This finding contrasts with the average patient group, which generally has an equal number of *FC* to $CF + C$ responses (Exner, 1993). Pure *C* responses are also predominant in the protocols of very young children, as is color naming (*Cn*; Ames et al., 1974; Exner, 1986). (See also interpretation of the $FC:CF + C$ formula.) It should be noted that there is strong evidence to interpret overall level of color used as a measure of how much emotions influence thoughts and behavior, though the evidence shows that Pure *C* responses should not be interpreted as extreme impulsivity or emotional reactivity (Mihura et al., 2013).

High Weighted Sum C

- Cognitive abilities have been suspended or possibly overwhelmed by affective impulses (possible aggressiveness and assaultive tendencies especially if combined with an absence of human movement).
- Possibly labile, suggestible, sensitive, and irritable.
- Difficulty delaying responses during problem-solving tasks.
- Color naming: indicates concrete, primitive, poorly conceptualized response to the stimuli (“stimulus bound”); this can reflect severe disorders for adults, such as organic impairment. (Note: Cn is not unusual in the protocols of young children, particularly if they have intellectual disabilities.)

Low Weighted Sum C

- Low spontaneity, overcontrolled emotions (e.g., depression, psychosomatic patients).
- Possible suicidal tendencies if other indicators are present.

High FC ($M = 2.97$, $SD = 1.78$)

- Good integration between controlling emotions and appropriately expressing them (moderate FC).
- Good rapport with others (moderate FC).
- Low level of anxiety, capacity to learn under stress, good prognosis for therapy (can conceptualize emotions and give form to their expression; moderate FC).
- May reflect overcompliance and a dependent personality (extremely high FC).
- Among children, may reflect the effects of overtraining with a corresponding decrease in natural spontaneity.

Low FC

- Poor emotional control, difficulties with interpersonal relationships due to the poor control.
- Possible anxiety states.
- Supports hypothesis of schizophrenia if other indicators of schizophrenia are present (i.e., high F– responses).

Color Achromatic (C' , $C'F$, FC')

Achromatic color (C') responses reflect constrained, internal, and painful affects. There is an irritation and a dampened emotional expressiveness in which the person is cautious and defensive. Exner (1993) referred to C' as the psychological equivalent of “biting one’s tongue, whereby emotion is internalized and consequently creates some irritation” (p. 386). Thus, it relates not only to painful emotions but also to affective constraint and defensiveness. Most Rorschach systematizers have consistently used C' as an index of depression. In considering the meaning of achromatic color responses, a clinician should look at the relative influence of form. If form is dominant (FC'), there is likely to be definition and organization to the affect, with a stronger ability to

delay the behavior. Dominant C' responses suggest the immediate presence of painful emotions. The average number of achromatic color responses for nonpatients is 1.49 (Exner, 2003). In contrast, individuals with depression have an average of 2.16, with individuals with character disorders averaging .83 (Exner, 1993). There is support for interpreting C' in this manner (Mihura et al., 2013).

High C' ($M = 1.60$, $SD = 1.33$)

- Highly constrained and painful emotions (e.g., people with psychosomatic, obsessive-compulsive, depressive disorders).
- Poor overall adjustment.
- Possible suicidality (with absence of shading responses combined with a large proportion of C' responses).

Shading—Texture (T , TF , FT)

Texture responses represent painful emotional experiences combined with needs for supportive interpersonal relationships (S. J. Beck, 1945, 1968; B. Klopfer et al., 1956). For example, recently divorced or separated subjects averaged 3.57 texture responses per protocol ($SD = 1.21$), as compared with 1.31 for matched controls ($SD = 0.96$; Exner & Bryant, 1974). Persons with a high number of texture responses need interpersonal connection and reach out, although they may do so in a guarded and cautious manner (Hertz, 1976). If form plays a relatively insignificant role and texture is predominant, individuals tend to feel overwhelmed by painful experiences, which would probably be sufficiently intense to disrupt their ability to adapt. Conversely, if form dominates (FT), not only is the pain likely to be more controlled, but also the need for supportive contact from others would be of primary concern (S. J. Beck, 1968; B. Klopfer et al., 1956). Coan (1956) suggested that a combination of movement and texture responses relates to inner sensitivity and empathy. If chromatic color and texture occur together, the person's behaviors would not only be less mature in seeking affection but also more direct and unconstrained (Exner, 1974). Support has been found for interpreting this variable in this way for high T , though less for low T (Mihura et al., 2013).

Responses in which texture dominates show an increase through childhood, reach a maximum by 15 years of age, and gradually subside over the next few years until form-dominated texture responses are most characteristic in late adolescence and adulthood (Ames et al., 1974). Nonpatient populations average 1 texture response per record, whereas psychiatric populations average 2 or more per record (Exner, 2003; Exner & Erdberg, 2005). Texture responses appear 10 times more frequently on Cards IV and VI than on the other cards (Exner, 1993).

High T or TF (for T only $M = 1.01$, $SD = 0.69$)

- Intense needs for affection and dependency.
- Oversensitivity in personal relationships, difficulty reconciling the intensity of needs for support with what they can realistically expect.
- Open to environment but approach it with a cautious sensitivity.

Low (absent) T

- Emotional “impoverishment,” person has ceased to look for meaningful emotional relationships (e.g., inpatients with depression, patients with psychosomatic disorders).
- Constrained expression of affect.

Shading—Dimension (Vista; V, VF, FV)

Rorschach systematizers have generally considered Vista responses, especially pure V, to represent a painful process of self-examination in which the person creates a sense of distance from self (Exner, 1993). This introspection usually involves depression, negative self-evaluation, and a sense of inferiority. However, if the V responses are dominated by form, introspection is still suggested, but the process is unlikely to be quite as negative and emotionally painful. This is in contrast to the negative type of self-examination associated with pure V. Even a single pure V response in a Rorschach protocol can be an important indicator. Research has supported this interpretation of V, but the effect size is quite small (Mihura et al., 2013). Practitioners should be cautious in the emphasis placed on this variable.

In normal populations, V responses occur, on average, 0.35 times per record (Exner & Erdberg, 2005). Depressed inpatients average 1.09, and persons with schizophrenia and character disorders average 0.60 and 0.24, respectively (Exner, 1993, 2003). It is extremely rare for V to appear in the protocols of children, but it occurs at about the same rate among adolescents as it does for adults (Exner, 1993; Exner & Weiner, 1995).

High V ($M = 0.35$, $SD = 0.77$)

- Deep self-critical introspection.
- Possible suicidal risk (see Exner’s Suicidal Constellation composed of 12 possible signs including high V, high number of morbid responses, $es > EA$, etc.; cutoff of 8 or more to identify high suicidal risk).

Low V

- Absence of V is a positive sign.
- Presence of a single form-dominated V represents the ability to introspect, suggests the resulting information can be integrated and eventually used productively.

Shading—Diffuse (Y, YF, FY)

B. Klopfer et al. (1956) and S. J. Beck (1945) described Y as representing a sense of helplessness and withdrawal, which is frequently accompanied by anxiety and is often a response to ambiguity. Beck further elaborated that persons with a high number of Y responses are experiencing psychological pain and have resigned themselves to their situation. Y increases during stress, such as before examinations (Ridgeway & Exner, 1980) and surgery (Exner, Thomas, Cohen, Ridgeway, & Cooper, 1981), with uncontrollable laboratory-induced stress (McCown et al., 1992), and during situational crises (Exner, 1993, 2003). Good support has been found for this variable (Mihura et al., 2013).

The same general rule for looking at the influences of form (F) in relation to Vista (V), texture (T), and color (C, C') also applies for shading-diffuse. When F is dominant, individuals are more able to delay their behavior, and their experience is more controlled, organized, and integrated. This ability to delay behavior also gives them time to mobilize their resources. When Y is dominant, there is a much greater sense of being overwhelmed. Although these individuals are characteristically withdrawn, any expression of pain and helplessness is direct. Because there is little ability to delay their impulses, they do not have enough time to mobilize their resources.

In the general population, 86% of people give at least one Y (Y, YF, or FY) response. Exner's (1993, 2003) normative groups of adult nonpatients had an average of 0.97 Y responses ($SD = 1.20$; Exner & Erdberg, 2005), compared with 2.12 for people with schizophrenia ($SD = 2.62$), and 1.81 for depressives ($SD = 1.40$). The total absence of Y suggests an extremely indifferent attitude toward ambiguity (Exner, 1993). To accurately understand the meaning of Y responses, the clinician should look for other indicators of coping. In particular, these might include the number and manner in which pure form is used, the quality of organization, and the number of human movement responses. If there is a high number of Y and these "coping indicators" are absent, the person is likely to be overwhelmed and probably unable to adapt or respond effectively (Exner, 1993).

High Y ($M = 0.97$, $SD = 1.20$)

- Anxiety, constrained expression of emotions.
- Resignation to life events, attempt to create distance between oneself and the environment.

Form Dimension

Form Dimension (FD) was included in the Comprehensive System because it seemed to be both an empirically distinct category and a source of some interpretive significance. Some research that exists suggests that a high number of FD responses are related to introspection and self-awareness. For example, a relatively high number of FD responses have been found among persons who are introverted, those who are involved in the later phases of insight-oriented therapy, and patients who have completed a wide number of other forms of therapy (Exner, 1993). FD responses occur more frequently among nonpatients ($M = 1.43$; Exner & Erdberg, 2005) than among other patient groups—including individuals with schizophrenia ($M = 0.60$) and depression ($M = 0.82$), and is particularly low among individuals with character disorders ($M = 0.33$; Exner, 1993, 2003). Most research, however, does not support this variable (Mihura et al., 2013). As such, practitioners should strongly consider not interpreting FD.

High FD ($M = 1.43$, $SD = 1.15$)

- Introspection.
- Self-aware, able to delay and internalize behaviors.

Pairs (2) and Reflections (rF and Fr)

Research on pairs and reflections has been linked both conceptually and empirically to self-absorption (Exner, 1991, 1993, 2003). However, this does not necessarily mean

that an individual with high reflections and pairs is pathological. For example, a high number of reflections was found among nonpatients in occupations that encourage a high level of self-worth, such as performing artists and surgeons (Exner, 1993). Whereas reflections occurred in only 7% of adult outpatients, they occurred in a full 20% of the protocols of those with character disorders and 75% of the records of antisocial groups (Exner, 1993). It is fairly common for children between the ages of 5 and 10 to have a high number of reflection (and pair) responses, but such responses usually decrease by adolescence, when individuals move to a less egocentric style of functioning (Exner, 1993; Exner & Weiner, 1995). There is support for this interpretation of these variables, especially reflections (Mihura et al., 2013).

High Pairs (2) and Reflections (rF and Fr) ($Fr + rF$, $M = 0.20$, $SD = 0.67$)

- Possible self-absorption.
- Inflated sense of self-worth, exaggerated sense of self-pride, with strong strivings toward status.
- Narcissistic tendencies.
- Need for self-affirmation may cause affective or interpersonal difficulties if they do not receive external validation.

Organizational Activity (Z)

The relative extent to which a person efficiently and effectively organizes the disparate aspects of the inkblots and sustains his or her cognitive efforts will be reflected in the scoring for Organizational Activity. The possibility that Organizational Activity is conceptually related to intelligence is given some empirical support in that moderate correlations (.54) have been found with the Wechsler-Bellevue Full Scale IQ and an even higher correlation of .61 exists with the Wechsler Vocabulary subtest (see Exner, 1993). Adults and younger nonpatients will have frequencies of Organizational Activity (Zf) averaging 13.5 and ranging between 9 and 17.5 (Exner & Erdberg, 2005). Among psychiatric patients, lower organizational activity has been noted among depressed patients (Hertz, 1948). In contrast, quite high levels of organizational activity have been found among patients who projected organized delusions (S. J. Beck, 1945; see also the interpretation for Processing Efficiency in the “Processing Section” topic later in the chapter). There has been support for interpreting this variable, especially the frequency of Z (Zf), in this way (Mihura et al., 2013).

High Zf (>13) ($M = 13.45$, $SD = 4.22$)

- High intellectual striving.
- Careful, precise work with perceptions.

Low Zf (<9)

- Person expends less effort than needed or required to adequately process information.

Content

The different content categories are generally considered to contain information relating to a person’s needs, interests, preoccupations, and social interactions.

Positive correlations have been found between a large variety of contents and intelligence (Exner, 1986). Research has also shown that whereas a high variety of contents is associated with intellectual flexibility, a low variety suggests intellectual constriction and rigidity. People's occupational interests are often represented in a higher number of contents relating to their specific career choices. For example, biologists and medical personnel usually give a higher number of anatomy responses than the general population (Exner, 1974). This finding may indicate that these persons merely have an interest in their career or that images related to their career are more familiar and thus more accessible in the mind, but it could suggest they are overconcerned with their career to the extent that they neglect other areas of their lives, perhaps even impairing their overall level of adjustment. For example, biologists who see only nature contents may be using a preoccupation with their careers to withdraw from interpersonal relationships (Exner, 1974), though this interpretation has not had much research support.

When interpreting Rorschach content, it is important to look at the variety of contents, the number of each content, and their overall configuration, as well as the implications other Rorschach factors may have for the meaning of the content scorings. It is usually essential to consider the age of the respondent and to use age-appropriate norms. For example, children usually have significantly fewer human and human detail responses than adults, and the variety of their contents is also lower (Ames et al., 1974; Exner & Weiner, 1995). Another important step is to study contents relating to aggressiveness (fire, explosions, etc.), facial features, and orality. Although the focus of the Comprehensive System is on a quantitative approach to the Rorschach, symbolic considerations can also be extremely important in conducting a more qualitative analysis. The next section provides general information on the meaning associated with human and animal contents. Further interpretive material can be found in the interpretation of quantitative formulas relating to contents, such as the Intellectualization Index, Isolation Index, Interpersonal Index, $(H) + (Hd):(A) + (Ad)$, and $H + A:Hd + Ad$.

Human and Human Detail [H, Hd, (H), (Hd)]

Human responses constitute one of the most thoroughly researched contents. S. J. Beck (1961), in general agreement with other researchers, has found that H and Hd gradually increase with age until the median for 10-year-old children is from 16% to 18%. This remains unchanged through adolescence, and the overall adult proportion of 17% is eventually reached. Exner (1974) found that, whereas adult nonpatient $H + Hd$ responses were 19% of the total adult outpatients' responses, persons with schizophrenia had a lower total of 13%. He also demonstrated that the ratio of human to human detail ($H:Hd$) for nonpatients was 3:1. In contrast, the average ratio for those with schizophrenia was approximately 1:1, and outpatients' ratio was 2:1. Molish (1967) suggested that when there is an increase in Hd compared with H, the person is prone to use constricted defenses. Others have theorized that the increase suggests intellectualization, compulsiveness, and a preoccupation with the self that restricts the degree of contact with others (B. Klopfer & Davidson, 1962). S. J. Beck (1945) associated high Hd with anxiety, depression, and a low intellectual level (see also the quantitative formula for Interpersonal Index). There is support for whole, realistic human content representing individuals viewing themselves and others as whole, integrated people (Mihura et al., 2013).

High H ($M = 3.18$, $SD = 1.70$)

- Wide-ranging interest in people.
- Possibly good self-esteem and high intelligence.
- Greater likelihood of successful psychotherapeutic treatment.

Low H

- Low level of empathy.
- Withdrawal from interpersonal relationships (i.e., low among people with schizophrenia).
- Poor prognosis for psychotherapy (with abrupt termination if M responses are low).

Animal and Animal Detail (A and Ad)

Most of the literature indicates that animal content is associated with the obvious aspects of adaptiveness and the most concrete features of reality testing (Draguns, Haley, & Phillips, 1967). Because animal contents are the easiest to perceive, their presence suggests that examinees are using routine and predictable ways of responding. Conversely, a low number of animal responses suggests highly individualistic persons who see their world in their own personal and unique ways, though these interpretations have not been thoroughly supported in the research literature.

Animal responses occur more frequently than any other content category with an average of 8.2 ($SD = 2.56$) for nonpatient adults (Exner & Erdberg, 2005), and a slightly higher frequency for children (S. J. Beck, 1961). Individuals with schizophrenia average 31%, whereas those with depression score much higher, averaging 41% per protocol (Exner, 1974). Other studies have found that the percentage of A responses is low for people with manic states (Kühn, 1963; H. Schmidt & Fonda, 1954) and high for people with alcoholism (Buhler & LeFever, 1947).

High A ($M = 8.18$, $SD = 2.56$)

- Predictable, stereotyped manner of approaching the world.
- Often associated with depression and the use of constrictive and conforming defenses.

Low A

- Persons who are spontaneous, nonconforming, unpredictable, and of higher intelligence often have a low number of A responses.

Anatomy (An) and X Ray (Xy)

Because An and Xy both measure concern with the body, they are considered together. Anatomy (An) responses have been well researched, and, along with human and animal contents, anatomy is one of the most frequently occurring responses (average of 0.6 for nonpatient adults). Anatomy content has an obvious connection with concern for the body, and the literature supports this connection in that it occurs more frequently for persons preparing to undergo elective surgery (Exner, Armbruster,

Walker, & Cooper, 1975) and among patients with psychosomatic disorders (Shatin, 1952). Anatomy responses also occur with greater frequency with the onset of psychological difficulties related to pregnancy (Zolliker, 1943). A review of the literature by Draguns et al. (1967) concluded that anatomy content can serve as an index of the degree of involvement persons have in their inner fantasy life or may reflect physical changes, such as illness, puberty, or pregnancy. Exner (1993) also suggested that anatomy content is associated with withdrawal from the environment and obsessive defenses. However, the research support for An and Xy centers on preoccupation on the functioning and vulnerability of one's body (Mihura et al., 2013).

The relative proportion of An to Xy responses may be an important consideration. Although anatomy responses are generally low for both psychiatric and nonpsychiatric groups, a combined An and Xy score allows for a clearer differentiation between the two groups. Whereas the combined An and Xy responses for a nonpatient group give a combined average of only 0.96 responses (Exner & Erdberg, 2005), outpatients give 1.5, people with schizophrenia give 1.4, and patients without schizophrenia give 1.8 (which accounts for 9% of this last group's total number of responses; Exner, 1974, 2003). Xy responses have been found to be particularly high for people with schizophrenia with bodily delusions (average of 2.2) and depressed patients with concerns related to bodily functioning (1.7; Exner, Murillo, & Sternklar, 1979). Anatomy responses occur most frequently for Cards VIII and IX, and Xy responses are most frequent for Card I. Research generally supports these interpretations of the combined An and Xy codes (Mihura et al., 2013) (Anatomy $M = 0.88$, $SD = 1.05$; X Ray $M = 0.08$, $SD = 0.28$).

Food (Fd)

A high number (2 or more) of food contents (primary or secondary) suggests dependency. High scorers would be expected to request extensive help and guidance from others, have difficulty making independent decisions, and be naive in their expectations of others (Exner, 1993). While there is research support for this interpretation, the effect size is low (Mihura et al., 2013), and practitioners should consider using the Oral Dependent Language (ODL%) Scale (R. F. Bornstein, 1996; 1999; R. F. Bornstein & Masling, 2005; Meyer, 2004) as a stronger measure of dependent tendencies ($M = 0.26$, $SD = 0.55$).

Popular Responses

The number of Popular responses reflects the respondents' degree of similarity to most people (especially the way they perceive the world), the extent to which they conform to social standards, and the relative ease with which they can be influenced in interpersonal relationships. Persons who reject conventional modes of thinking give a significantly lower number of Popular responses than those who are conforming and relatively conventional. Good support has been found for these interpretations (Hiller et al., 1999; Mihura et al., 2013). With Exner's (2003) scoring system, the average number of P responses for nonpatients is 6.28 ($SD = 1.53$; Exner & Erdberg, 2005). Outpatients and patients without schizophrenia, likewise, give approximately 7 per record, whereas inpatients with schizophrenia give 4 or less, persons with a character

disorder give approximately 5, and people with depression have slightly more than 5 (Exner, 1993).

Because Popular responses are extremely common for Cards I, III, V, and VIII, an absence of them from these cards is significant in that it more strongly suggests the trends just discussed. However, the assumption that low P responses alone confirm maladjustment should be approached with caution. Low P subjects who have good form quality (F + % and X + %) and whose organizational activity is also good are likely to be creative individuals who are avoiding common or ordinary perceptions and want to extend their imagination. If organization and form quality are poor, there is a high likelihood that the psychopathological dimensions are more predominant.

High P ($M = 6.28$, $SD = 1.53$; high considered > 7 , low considered < 4)

- Conventional, overconforming, guarded, and frequently depressed.
- Anxiety related to a fear of making mistakes and, therefore, clings to common perceptions as a way to achieve approval.

Low P

- Poorly adjusted, detached, aloof from their environment, and unable to see the world as others see it.
- Possible character disorder reflecting their rejection of conventionality and their lack of conformity.
- Highly creative (if F+%, X+%, and organizational activity are high).

Special Scores

Deviant Verbalization (DV), Deviant Responses (DR), Incongruous Combination (INCOM), Fabulized Combination (FABCOM), Contamination (CONTAM), Inappropriate Logic (ALOG)

The first six of the Special Scores were included in the Comprehensive System to detect the presence of cognitive slippage. Illogical, dissociated, fluid, or circumstantial thinking is particularly likely if there are Level 2 codings for the first four Special Scores (Exner, 1991). This is consistent with the finding that virtually no Level 2 DV or DR responses occurred among nonpatients, but an average 1.90 Level 2 DRs have been noted among those with schizophrenia (Exner, 1993). However, there is no specific interpretation for each of the six categories. Instead, they are used collectively to detect the presence and seriousness of cognitive distortions. The relative seriousness is indicated in part by the type of Special Score. Mild distortions are suggested by the presence of scores for DV (Level 1), INCOM (Level 1), or DR (Level 1), and moderate distortions are suggested by the presence of DV (Level 2), FABCOM (Level 1), INCOM (Level 2), and ALOG. The most serious degree of cognitive distortion is suggested if respondents have Special Scores for DR (Level 2), FABCOM (Level 2), and CONTAM. There is good research support for interpreting Level 2 Special Scores as thought disturbance (Mihura et al., 2013) (Deviant Verbalization $M = 0.34$, $SD = 0.67$; Deviant Response $M = 0.85$, $SD = 1.01$; Incongruous Combination $M = 0.71$, $SD = 0.93$; Fabulized Combination $M = 0.45$, $SD = 0.77$; Contamination $M = 0.00$, $SD = 0.00$; Inappropriate Logic $M = 0.04$, $SD = 0.21$).

A further means of analyzing the first six Special Scores is by noting the relative elevation of WSum6, which is simply a sum of the different weightings given to the Special Scores. The weightings are as follows:

- Level 1 Deviant Verbalization (DV1) = 1
- Level 2 Deviant Verbalization (DV2) = 2
- Level 1 Deviant Response (DR1) = 3
- Level 2 Deviant Response (DR2) = 6
- Level 1 Incongruous Combination (INCOM1) = 2
- Level 2 Incongruous Combination (INCOM2) = 4
- Level 1 Fabulized Combination (FABCOM1) = 4
- Level 2 Fabulized Combination (FABCOM2) = 7
- Inappropriate Logic (ALOG) = 5
- Contamination (CONTAM) = 7

A weighted score is given each time the code is given. For example, three occurrences of a Level 1 Deviant Response (DR) would equal a sum weighted score of 9. The average WSum6 for nonpatients is 7.2 (Exner & Erdberg, 2005), indicating that normals do generally include at least some of the Special Score responses. In striking contrast are individuals with schizophrenia, who have an average WSum6 of nearly 45 (Exner, 1993). However, the presence of Special Scores does occur among children under 10 but gradually decreases during adolescence (Exner & Weiner, 1995). The general interpretation for a high sum of weighted Special Scores is that there is cognitive distortion and thought disturbance. The interpretive hypothesis, especially with a high WSum6, is that there is a serious disregard for reality, strained reasoning, faulty cause-and-effect relationships, loose associations, disorganized thinking, and poor ability to focus on tasks (Exner, 1991, 1993). This ability of the Rorschach to detect the bizarre and illogical processes of schizophrenia is probably one of its best-validated features (Hiller et al., 1999; Mihura et al., 2013; Vincent & Harman, 1991), and there is some evidence that it is sensitive to these changes in thought processes even before their clinical manifestation (G. Frank, 1990) (WSum6 $M = 7.12$, $SD = 5.74$).

Perseveration (PSV)

The presence of perseveration has been considered to represent some difficulty in cognitive shifting. Thus, the individual may have either a permanent or a temporary difficulty with rigidity or inflexibility in information processing or decision making (Exner, 1993). There is research support for interpreting this variable in this way (Mihura et al., 2013) ($M = 0.99$, $SD = 1.10$).

Abstract Content (AB)

The presence of one or more abstractions suggests intellectualizing defenses (see Intellectualization Index), although there is little research on the validity of this interpretation ($M = 0.21$, $SD = 0.56$).

Aggressive (AG) and Cooperative Movement (COP)

It is useful to consider AG and COP together. If there is an absence of scores in either category, it suggests that the individual is aloof, somewhat uncomfortable in social

situations, and on the periphery of group situations. In contrast, if COP is high (2 or more) and AG is low (0 or 1), the person is likely to be perceived by others as trustworthy, cooperative, and easy to be around (Exner, 1993). It is also a favorable prognosis for psychotherapy. If COP is low (<3 or especially = 0) and AG is high (> 2), the person's interactions are likely to be forceful or even aggressive and hostile (Exner, 1993). Given these interpretations, it might be speculated that high scores on both COP and AG would suggest some conflict regarding the appropriate and preferred mode of responding and would result in inconsistent interpersonal behaviors (e.g., passive aggressive interactions). Moderate support has been found for these variables, although it should be noted that, while both were found to be significant, the effect size for AG was low, while there is stronger evidence for these interpretations of COP (Mihura et al., 2013) (Aggressive $M = 0.89$, $SD = 1.02$; Cooperative Movement $M = 2.07$, $SD = 1.30$).

Morbid Content (MOR)

Although the presence of one MOR is not unusual in the records of nonpatients, two or more suggest pessimism, a negative self-image, and possible depression and is consistent with a diagnosis of posttraumatic stress disorder (Weiner, 1996). If three or more MOR responses are present, it is both a strong indicator of depression and one of several indicators for suicide risk (see the "Suicide Constellation" discussed later in the chapter; Exner, 1991, 1993). Research support has been found for this interpretation of this category (Mihura et al., 2013). MOR content is likely to have unique meaning for the person and can often be interpreted symbolically and qualitatively ($M = 0.93$, $SD = 1.01$).

Good Human Representation (GHR) and Poor Human Representation (PHR)

GHR and PHR are considered dichotomous categories. Persons with a high number of GHRs have a healthy, adaptive understanding of other people and are usually highly regarded by others, well adapted, competent, and reasonably free from chaos (Exner, 2003). In contrast, persons with psychiatric histories typically give a low number of GHRs. If they also give a high number of PHRs, they have distorted understandings of other people and are also likely to report histories of interpersonal difficulties, are socially inept, and are interpersonally ineffective (Exner, 2003). Research has provided support for these interpretations, especially for GHR (Mihura et al., 2013) (Good Human Representation $M = 5.06$, $SD = 2.09$; Poor Human Representation $M = 2.12$, $SD = 1.81$).

Personal (PER)

Scores of 3 or more suggest a defensive authoritarian stance in which the individual is insecure regarding challenges to his or her sense of self. Interaction with the world is colored by justification of values, attitudes, and behaviors based on the person's own personal experience. Interpersonal difficulties may be experienced during attempts to get others to submit to his or her opinions (Exner, 2003). There has been research support for these interpretations (Mihura et al., 2013) ($M = 0.99$, $SD = 1.10$).

Color Projection (CP)

This highly unusual response indicates persons who deny unpleasant emotions by creating false or substitute emotions instead. Thus, they have difficulty dealing with negative feelings and modulating their emotions, and they bend or even distort reality as a means of adapting (Exner, 1993, 2003). This scoring category should be interpreted only in the context of other indicators for processing and expressing affect (see the “Affect Section” later in the chapter). Research has not been conducted on this variable to support or refute these interpretations (Mihura et al., 2013) ($M = 0.01$, $SD = 0.11$).

Ratios, Percentages, Derivations

The quantitative formulas used to develop the different ratios, percentages, and derivations provide a more in-depth and complicated portrayal of the relationships among the Locations, Determinants, Contents, Populars, and Special Scores. These formulas provide some of the most important, reliable, and valid elements of interpretation. Their numbering and organization correspond with the numbers given to them in the previous listing of the quantitative formulas (see the section titled “Comprehensive System: Scoring the Structural Summary”).

Core Section

The Core section provides information on the person’s dominant personality style, particularly focusing on the level of stress the person is experiencing and how effectively he or she can tolerate the stress. Seven of the entries for the Core section are frequency data providing summaries for total number of responses (R), animal movement (FM), inanimate movement (m), Achromatic color (C’), Shading—Texture (T), Shading—Dimension (V), and Shading—Diffuse (Y). Interpretive material for each of the last six categories can be found in previous sections; the first category, number of responses (R), is detailed in the subsection that follows.

Number of Responses Number of responses is not a quantitative formula (and is, therefore, not numbered). Rather, it is a simple sum of the total number of responses. In using Exner’s set of instructions, the mean for the total number of responses for nonpatient adults is 23.36 ($SD = 5.68$; Exner & Erdberg, 2005). However, different methods of administration can influence this number to a certain extent. For example, Ames et al. (1973) reported an overall adult average of 26; S. J. Beck (1961) gave 32 for his adult mean; and both use instructions somewhat different from Exner’s. Deviations from the normal range present the following possible interpretive hypotheses. There is good support for interpretations based on R (Mihura et al., 2013).

Low R (Adults, < 17; Children, < 15) ($M = 23.36$, $SD = 5.68$)

- Defensiveness (possibly consistent with malingering).
- Constriction, depression.
- Invalid profile (is less than 14; see administration instructions).

High R (> 27)

- Introversion.
- Above-average intelligence with a relatively high level of academic achievement, high degree of creativity.
- Good ego functioning, including the ability to plan ahead, adequate impulse control, and the ability to tolerate stress.
- Among patient populations: mania, obsessive-compulsive disorder.
- Invalidates formulas (higher proportion of D and Dd, more pure F, more Pop-ulars, elevated Affective ratio due to greater number of responses to Cards VIII and X).

1. Lambda (L). The Lambda index was developed by S. Beck (1961) as an improvement on the F% that had been used by other Rorschach systematizers. The earlier F% used the total number of R as the denominator, whereas the Lambda uses the total number of nonpure F.

$$\frac{F \text{ (number of responses having only Pure F determinants)}}{R - F \text{ (total R minus Pure Form answers)}}$$

In calculating Lambda, only responses involving form are used (F, M, CF, etc.) and not determinants without form (C, C', T, etc.). The Lambda ratio is used as an overall index of the degree of responsiveness versus lack of responsiveness to stimuli (Exner, 2003). This includes how attentive individuals are to complexity and nuance in the world. Thus, persons can range from highly constricted and withdrawn to completely emotionally flooded by their responses to stimuli. The Lambda for nonpatients is between 0.11 and 2.33, with a mean of .58. In contrast to this is the much higher range among persons with schizophrenia (.05–29.00), depression (.08–15.00), and character disorders (0.015–16.00; Exner, 1993, 2003). This greater range among patients reflects their greater tendencies either to overreact to stimuli or, in contrast, to underreact by becoming highly constricted and withdrawn. Thus, a maladjusted person may have a Lambda either greater than 0.99 (avoidant) or less than 0.32 (overly engaged). The significance lies in Lambda's ability to provide specifics regarding the form this maladjustment takes. It is also important to look at other information within the test, such as form quality and Experience Balance, to obtain a more complete conceptualization of the meaning of L. However, with adolescents, an interpretation that focuses on maladjustment should be made with caution because adolescents usually have a higher proportion of pure F responses (Ames et al., 1974; Exner, 1995). There is evidence to support the interpretation of Lambda as presented here (Mihura et al., 2013).

High L ($M = 0.58$, $SD = 0.37$; high L > .99)

- Withdrawal from experiencing a situation fully, avoidance of perceiving all the possibilities that may be present ("tunnel vision").
- Likely to be conservative, insecure, detached, and fearful of involvement.
- Defensive, constricted, unimaginative, anxious.
- Possible depression, guilt, increased potential for suicide.

Low L (low L < 32)

- Overinvolvement with stimuli to the extent that affect disrupts cognitive functioning.
- Inadequate control over emotions; frequent, impulsive acting-out results in difficulty maintaining satisfactory interpersonal relationships.
- Impaired ability to attend to their environment, victims of their needs and conflicts.
- Achievement-oriented persons who deal effectively with their environment (if these indicators in their protocols reflecting control and flexibility are present: average X + %, average number of Populars, good Organizational Activity, above-average W).

2. Experience Balance, or Erlebnistypus (EB). The Experience Balance formula, or Erlebnistypus, was originally devised by Rorschach and is the ratio between the sum of all M responses compared with the sum of all weighted color responses. The ratio is expressed as Sum M: the Weighted Sum Color (WSumC). The Weighted Sum Color side of the ratio is calculated according to this formula:

$$WsumC = (0.5) \times FC + (1.0) \times CF + (1.5) \times C$$

All human movement responses are included in the formula, regardless of whether they are the major determinant of the response. Color naming responses are not included.

Rorschach systematizers and researchers have come to view the Experience Balance ratio as the extent to which a person is internally oriented as opposed to being more externally directed and behaviorally responsive to outside stimuli. Although the EB ratio is usually relatively stable (Exner, Armbruster, & Viglione, 1978), it can temporarily change during times of stress or become more permanently altered during the course of successful psychotherapy (Exner, 1974; Exner & Sanglade, 1992). Although the EB ratio is usually stable for adults, there is considerable variability in children until midadolescence (Exner et al., 1985; Exner & Weiner, 1995). In an extensive literature review, J. Singer (1960) described the two sides of the ratio as representing dimensions of “constitutional temperament.” These dimensions are introversives (higher M scores), who have a preference for internal experience, as opposed to extratensives (higher weighted C scores), who are more prone to activity and external expression. An introversive can more effectively delay his or her behavior, whereas the extratensive is more emotional and is likely to discharge his or her affect into some form of external behavior. Both types respond differently to stress and to problem-solving tasks (Exner, 1978). It should be emphasized that, in their moderate forms, neither is any more or any less effective than the other, nor is either more prone to psychopathology (Molish, 1967; see also the interpretive meanings associated with M and C and the quantitative formulas dealing with either of these factors [EA, EBPer, D score, Adjusted D score, and W:M]). Despite this research, the preponderance of studies has not supported interpretation for this index (Mihura et al., 2013), so interpretations should be made with caution.

Higher M (Introversives)

- Oriented toward using their inner fantasy life, directed inward, and use their inner experience to satisfy most of their basic needs (even though externally they may have learned to appear to be extraverted).
- Cautious, deliberate, submissive.
- Less physically active than persons scoring relatively higher on the C side of the ratio.
- Approach problem-solving tasks by internalizing the situation and mentally reviewing possible alternatives.

Higher C (Extratensives)

- Use external interactions as the most important means of satisfying their needs.
- Difficulty delaying their responses.
- Direct their energy toward the outside world.
- Spontaneous and assertive.
- Approach problem-solving situations by experimenting with different behaviors (external trial and error) before achieving solutions.
- Among children, may represent a lack of self-assurance.

M and C Roughly Equal (Ambitents)

- Flexible during interpersonal relationships.
- Less sure of themselves during problem solving and tend to vacillate, usually need to verify every sequence in the solution of a problem at hand, do not profit as much from mistakes as either introversives or extratensives.
- Among patient populations, unusually high scores on both M and C suggest a manic condition.

3. Experience Actual (EA). Whereas the Experience Balance ratio emphasizes the assessment of a person's type, the Experience Actual ratio indicates the "volume of organized activity" (S. J. Beck, 1960). That is, it is a measure of the overall resources an individual has to cope with the world. It is calculated as:

$$\text{Sum of Human Movement} + \text{Weighted Sum Color}$$

Although the M side of the formula shows the extent to which persons are able to organize their inner lives and the C side indicates the extent to which emotions are available, the emphasis here is that both the M and the C represent deliberate, organized activity to engage with and adapt to the world. This is contrasted with the disorganization associated with nonhuman movement (FM, m) and the responses related to the grey-black features of the blot (T, V, Y).

For the most part, the adult ratio between M and C is remarkably stable (Exner, 1993), yet the sum of M and C sometimes fluctuates on a daily basis, which theoretically parallels the effects of changes in mood (Erginel, 1972). After successful psychotherapy, particularly if long term, M and C typically both increase (Exner & Sanglade, 1992; Weiner & Exner, 1991), indicating a greater increase in the degree of organization

of the person's inner life and an availability of more emotions. In fact, Exner (1974) found that EA increases significantly more for patients who improved in therapy than for those who showed little or no improvement. Furthermore, persons who underwent long-term, insight-oriented treatment showed much more of an increase in EA than those in a treatment that emphasized a combination of support and environmental manipulation (Exner, 1974). This is consistent with the goal of insight therapy, which focuses on helping patients to understand and organize their internal resources. The mean changes for children show a gradual increase (rarely more than 0.5) with each year from the ages of 5 to 13 (Exner, 1993). Although brief retesting for children has shown good stability, long-term retesting (9 months or more) has resulted in wide fluctuations (Exner et al., 1985; Exner & Weiner, 1995). Good support has been found for interpretations based on this index (Mihura et al., 2013) ($M = 9.37$, $SD = 3.00$).

4. Experience Pervasive (EBPer). Because Experience Balance (M:WSumC) is a somewhat crude indicator of how pervasive or dominant the introversive or extratensive style is, Experience Pervasive was designed as a more refined means of indicating how dominant one of the two styles is. Thus, it is an extension of the interpretations described in Experience Balance. It is calculated by dividing the larger number in the EB ratio by the smaller one. This is done only when a marked difference (style) is evident in the EB ratio. According to Exner (2003), it is calculated only when "the value of EA falls between 4.0 and 10.0 [and] one side of the EB [is] at least 2 points greater than the other side. If the value of EA is more than 10.0, one side of the EB must be at least 2.5 points greater than the other" (p. 237). Thus, it is calculated only when a clear style is indicated. When this occurs, it clearly indicates that one of the styles is quite pervasive, perhaps to the point of suggesting rigidity in problem-solving style (Exner, 1993). However, there is currently no research support for interpretations based on this index (Mihura et al., 2013).

5. Experience Base (eb). The Experience Base ratio was originally suggested by B. Klopfer et al. (1956) and later developed in its present form by Exner (1974, 1986). The Experience Base ratio compares all nonhuman movement determinants (FM + m) with the sum of all the shading and achromatic color determinants. It is summarized by this ratio:

$$\text{Sum FM} + \text{m} : \text{Sum C}' + \text{Sum T} + \text{Sum Y} + \text{Sum V}$$

The nonhuman movement side of the ratio reflects tendencies to respond in ways that are not completely acceptable to the ego. These tendencies appear out of control, impinge on the individual, and are disorganized (B. Klopfer & Davidson, 1962). Although the tendencies and feelings may have originally been produced by outside sources, the resulting internal activity is not in the person's control. The opposite side of the ratio, which is a sum of the responses relating to the grey-black features of the blot, is a reflection of the pain and disharmony the person is feeling as a result of unresolved stress. The eb ratio indicates which of these two areas of functioning is more predominant. If the eb is small on both sides, it suggests that the person is not experiencing very much pain and that his or her needs are well organized. Usually the values on either side of the ratio range between 1 and 3 for nonpatients. If either side becomes greater than 5, its interpretive meaning becomes more clear. Despite the conceptual and research findings, no clear research evidence has been found to support interpretation

for eb (Mihura et al., 2013). (See also the additional interpretive meanings associated with material from the left side of the ratio [FM and m] and the right side [Y, T, V, and C].)

6. Experienced Stimulation (es). Experienced Stimulation is the sum of the nonhuman movement responses and all responses relating to the grey-black features of the inkblot. This is accomplished by adding together both sides of the eb or by:

$$\text{Sum FM} + \text{m} + \text{Sum C'} + \text{Sum T} + \text{Sum Y} + \text{Sum V}$$

These are all responses reflecting that the person's functioning is disorganized and that forces are acting on him or her and the person feels those forces are beyond control. Thus, the es sum is an index of a person's degree of disorganization and helplessness. Persons scoring high on es have a low frustration tolerance, and it is difficult for them to be persistent, even in meaningful tasks (Exner, 1978). There is research support for interpreting this scale (Mihura et al., 2013).

Important information can be obtained by comparing the amount of organization the person has (as represented by EA) with how much chaos and helplessness he or she experiences (as represented by es). Normal populations usually have a higher EA than es, whereas psychiatric populations have a higher es than EA (Exner, 1974). Exner (1978) suggested that the ratio between EA and es can provide an index of the degree to which a person can tolerate frustration. Difficulty in dealing with frustration would primarily result from high-scoring es persons having a limited ability to process and mediate cognitive information (Wiener-Levy & Exner, 1981). As would be expected, a correlate of successful psychotherapy is that there is a decrease in es and a corresponding increase in EA, which suggests that at least some of the patient's activity has become more organized (Exner & Sanglade, 1992; Gerstle, Geary, Himmelstein, & Reller-Geary, 1988; Weiner & Exner, 1991). This finding was supported by Exner (1974), who found that individuals rated as unimproved after therapy also showed little change in their high es:EA ratio. In another study, Exner (1974) demonstrated that most persons in successful insight therapy had an increase in EA compared with es. This finding suggests that patients in successful insight therapy were able either to neutralize or to reorganize the forces that were "acting on" them. In contrast, therapy emphasizing support or environmental manipulation produced no or little change in the es:EA ratio.

High es ($M = 9.55$, $SD = 4.01$; high > 12)

- Low frustration tolerance.
- Difficulty following through on tasks.
- Disorganization, distractibility, and a sense of helplessness.

7. D Score (D; EA – es). The D score is a further measure of the client's ability to tolerate stress. It is essentially a means of evaluating the degree of available resources the person has (EA) versus the amount of disorganized events that are occurring beyond the person's control (es). It is calculated by first subtracting es from EA (EA – es) and designating whether the resulting number is a positive or negative number. The resulting raw score is then converted into a standard score by consulting a conversion table provided in Exner (2003; Table 10.4, p. 152). This number is a general measure

of coping abilities. For example, veterans diagnosed with posttraumatic stress disorder have been found to have low D scores (Weiner, 1996). Good research support has been found for this scoring category (Mihura et al., 2013).

Low D Score ($M = -0.12$, $SD = 0.99$; Low < -1)

- Person is likely to feel overwhelmed, overloaded, easily distracted, has limited psychological resources to deal with stress.
- Unable to deal with complex or ambiguous situations.
- Thoughts, affects, and behaviors might be impulsive and poorly focused; as the D score becomes progressively lower, this trend is likely to become increasingly stronger.

High D Score (High > 0)

- Client can adequately deal with the current level of stress.

8. Adjusted es (Adj es). Because es reflects general stressors and includes measures of current stimuli impinging on the person (m and Y), a different, adjusted es that excluded m and Y was developed. It is calculated by subtracting all but 1 m and 1 Y (including FY and YF) from es. It is believed that adjusted es represents the more chronic (rather than fluctuating) condition of the person (Exner, 1993). Thus, persons scoring high are likely to feel chronically overstimulated (e.g., racing thoughts, insomnia) and have difficulties organizing their thoughts. However, the main purpose of calculating Adjusted es is to enable the calculation of the Adjusted D score. Even though good support has been reported for D, little research supports interpretations for Adj es (Mihura et al., 2013).

9. Adjusted D score (Adj D). Because the D score evaluates levels of coping abilities but includes measures of current stress, it may not provide a measure of the person's usual ability to modulate and control his or her behavior. This issue is particularly likely to be present for clients referred for evaluation, because the events surrounding a referral usually involve psychosocial difficulties. These situational uncontrollable stressful events are expressed on the Rorschach (and in the D score) by the presence of m and Y responses (McCown et al., 1992). Adj es has had m and Y subtracted from it, so it theoretically removes the influence of current environmental stressors. Adjusted D is calculated by subtracting Adj es from EA, which produces a raw score that is converted to a standard score using the tables in Exner (2003; Table 10.4, p. 403). What remains in the Adjusted D score is a measure of the person's typical or usual capacity to tolerate stress and to control behaviors (Exner, 1993, 1995). Although there is some research support for interpreting this variable, the effect size is small (Mihura et al., 2013), so interpretations should be made with caution.

Low Adj D ($M = 0.19$, $SD = 0.83$; low < -1)

- Fewer than average resources to adequately cope with stressful situations.
- Function best in routine and predictable situations, adapting to new situations presents difficulties in that they are prone to become distracted, disorganized, and impulsive. (These trends are strengthened with progressively decreasing scores on Adj D.)

High Adj D (high > 1)

- Good ability to deal with stressful situations. (This does not *necessarily* mean that they are also well adjusted; antisocial and paranoid personalities have intricate systems of dealing with stress that are quite effective, but they are not well adjusted).
- May use their somewhat limited resources to distance themselves from the types of experiences that might result in increased growth and awareness, low motivational distress.

Ideation Section

The Ideation section focuses on information related to how the client imposes meaningful organization onto his or her perceptions. It includes three quantitative formulas (two ratios and an index) and frequency data for M–, M, number of Level 2 responses, WSum6, and M with no FQ (see the interpretation for each of these frequencies under the listings for “Human Movement and Weighted Color (MC)” and Special Scores).

1. Active:Passive Ratio (a:p). This ratio is calculated by adding the total number of active movement responses and comparing it with the total number of passive movement responses:

$$M^a + FM^a + m^a : M^p + FM^p + m^p$$

Individuals who have a distinctly higher number of passive responses are likely to be correspondingly more passive in other situations. In contrast, a clearly higher number of active responses indicates a person who is more active in terms of thoughts and behaviors (see also the interpretation for $M^a : M^p$ below). However, the contrast or magnitude of differences must be quite clear, as indicated by one of these conditions: (a) “sum of the values in the ratio is four and one value is zero”; (b) “values in the ratio exceed four, and the value on one side of the ratio is no more than twice that of the other”; or (c) “ratio exceeds four, and the value on one side is two to three times greater than the value on the other side” (Exner, 1993, p. 475). Practitioners should be cautious in interpreting this scale, as the research has not supported its interpretation (Mihura et al., 2013).

2. M Active:Passive Ratio ($M^a : M^p$). A further refinement of the a:p is to consider only the proportion of active and passive responses for human movement scorings. If the summed value of passive Ms (M^p) is greater than active Ms (M^a), it suggests a generally more passive orientation. For example, therapists’ ratings of clients with a greater number of passive Ms indicated that they made more requests for direction, seemed more helpless, and exhibited a relatively high number of silences (Exner, 1978). In addition, their daydreams had more passive themes (Exner, 1974), as did their Thematic Apperception Test (TAT) story endings (Exner, 1993). Despite this research and intuitive appeal, most research has not supported interpretations based on active/passive (Mihura et al., 2013).

3. Intellectualization Index . This index is calculated by multiplying the total number of Abstract (Ab) responses by 2, then adding the sum of Art and Ay responses according to this formula:

$$2Ab + (Art + Ay)$$

Earlier research indicated that the presence of three or more summed scores for Abstraction (Ab) and Art (Art) suggests an excessive use of intellectualization (Exner, 2003). Both individuals with obsessive disorders and paranoid schizophrenia were often found to have more than three combined Ab and Art frequencies in their protocols (Exner, 1986; Exner & Hillman, 1984), and both of these groups are likely to use an intellectual approach to distance themselves from their emotions. This is in contrast to nonpatients who typically reported an average score of approximately 2 per protocol (Exner & Erdberg, 2005). Despite these findings, very little research has looked at this scale (Mihura et al., 2013), so interpretations should be made with caution ($M = 2.17$, $SD = 2.15$).

High Intellectualization Index (5 or More)

- Neutralize emotions through analyzing things from an intellectual perspective, deny or conceal the impact of affect.
- Dealing with emotions is typically circumspect and possibly unrealistic.
- Intellectualization might provide people with a certain degree of control for moderate levels of affect, but much higher levels are likely to overwhelm them, quite possibly resulting in disorganization.

Affect Section

The Affect section provides information on how the person modulates and expresses affect. Because affect is expressed most directly on the Rorschach through color, the different frequencies and formulas are concerned with the various combinations of color with other types of Rorschach responses. Specifically, this section includes frequencies for Pure C, white space use (S), color projection (CP), and three quantitative formulas.

The sum of C and Cn responses provides an index of the degree to which a person is likely to be overwhelmed by affective impulses. Among nonpatient adults, it is rare to have any C or Cn responses occurring in a protocol ($M = 0.12$, $SD = 0.43$), but this increases slightly for patient groups (see discussion in the section on interpretation of color). The degree to which a person uses white spaces (S) has been associated with the person's negativism, means of handling anger, and amount of oppositional tendencies (see Interpretation section on white spaces). Color projection (CP), a rare response included as a Special Score, relates to a tendency for the individual to substitute alternative emotions in place of unacceptable unpleasant ones (in the Interpretation section, see the discussion of color projection [CP]). It is important to note the level of research support for each of these interpretive guidelines (see sections above).

1. Form-Color Ratio [FC:(CF + C)]. This ratio indicates the total number of form-dominated chromatic color responses, as compared with the absolute number of color-dominant chromatic responses. To calculate this formula, each of the chromatic color determinants is weighted equally as 1. Cn determinants are also included on the right side of the ratio because they are considered color-dominant responses. The ratio provides a measure of the degree of control a person has over his or her impulses (also check D score for a tendency to become overwhelmed by stress) and how emotionally

reactive he or she is. If form is predominant (1.5 to 2.5 times greater), it suggests the person has good control over his or her impulses and experiences satisfying interpersonal relationships (Exner, 1969, 1974; B. Klopfer & Davidson, 1962). Exner (1978), for example, found that people with schizophrenia who have FC responses greater than CF + C have a better response to psychotherapy and less likelihood of relapse. The high form suggests they can integrate an accurate, reality-oriented interpretation into their perceptions. However, if no or very few color-dominant responses (no CF + C) are present, the person will be overly constricted and have little contact with his or her emotions (Exner, 1978, 1993). This is consistent with the finding that most patients with psychosomatic disorders, who are typically emotionally constricted, had ratios of 4:1 or greater (Exner, 1993). If the CF + C side of the ratio is relatively high (1:1), it suggests a weak control over a person's impulses, which may be accompanied by impulsive behavior (Pantle et al., 1994) or aggressive acting out, perhaps consistent with a narcissistic personality (Exner, 1969; B. Klopfer & Davidson, 1962). The perception of both internal and external events is typically distorted and inaccurate, as are the responses to these events (Exner, 1974). The number of pure C responses increases with pathological groups, as indicated by only 7% of nonpatients giving pure C responses in contrast to 45% of those with depression, 32% of those with schizophrenia, and 27% of patients with a character disorder (Exner, 1993). Research has generally supported this index (Mihura et al., 2013).

2. Affective Ratio (Afr). Because the last three cards are chromatic and the first seven are primarily achromatic, the Affective Ratio indicates the extent to which affect (color) makes an impact on the person. It is composed of the total number of responses to the last three cards, compared with those given to the first seven cards, or:

$$\frac{\text{Number of responses to Cards VIII + IX + X}}{\text{Number of responses to Cards I + II + III + IV + V + VI + VII}}$$

Nonpatient adults usually show a mean Afr of .67 ($SD = 0.16$). However, it is relevant to consider Afr in the context of EB. Introversives (higher M side of EB), who primarily direct their experience inward, have Afr ranges between .50 and .80. In contrast, Extratensives (higher C side of EB) have Afr ranging between .60 and .95 (Exner, 1993, 2003). This means that it is useful to take EB scores into account when judging whether an Afr is high or low. Although the mean Afr for patient groups was not very different from that for nonpatients, the range was much higher for patients and the distribution was bimodal. This higher range among patient groups is consistent with the view that they are more likely to have difficulties with either undercontrolling or overcontrolling affect (Exner, 1993, 2003). However, while research has provided some support for this category, the effect size is relatively low (Mihura et al., 2013), suggesting that interpretations should be made only tentatively.

High Afr ($M = 0.61$, $SD = 0.17$; high > .85)

- Overresponsiveness to affect, person is more receptive to emotional inputs and more likely to respond immediately rather than delay behavior (check FC:(CF + C).

Low Afr (low < .53)

- Tend to withdraw from emotions; if extremely low, persons may attempt to exert an extreme amount of control over their affective responses (note the Intellectualization Index discussed earlier).

3. Complexity Index (Blends/R). Approximately 20% of all Rorschach responses involve blends, which are responses that include more than one determinant. To create a blend response, the person must appreciate the complexity of the inkblot, which requires both analysis and synthesis. Exner (2003) pointed out that the pure F response is the exact opposite of a blended response in that pure F requires attention to only the most simple, straightforward aspect of the stimulus. This index compares the total number of blend responses (entered on the left side of the ratio) with the total number of responses (R; entered on the right side of the ratio). Usually there are 1 or more blends in a person's protocol. A complete absence of blends suggests narrowness and constriction. This is consistent with the finding that blends are less frequent in the protocols of those with depression and persons with below-average intelligence (Exner, 1993, 2003). In contrast, an extremely high number of blends (8 or more) suggests an unusual amount of complexity, to the extent that the person may be overly burdened (Exner, 1993, 2003). Research has supported interpretation of this index (Mihura et al., 2013).

A thorough interpretation of blends also requires an understanding of their qualitative aspects. For example, a blend that includes color-dominated determinants implies that the person might be easily overwhelmed by affect, whereas the opposite would be true if the blend were form-dominated. The color-shading blend (combining color with C', Y, T, F, V) implies concern with painful, irritating, confusing emotional experiences, and it is associated with the protocols of depressed people. Exner and Wylie (1977) found a moderate correlation with attempted suicide. Accordingly, this type of blend was included as one of several variables in Exner's (1993) Suicide Constellation. However, the presence of color-shading blends does not seem to be a sufficiently accurate predictor of suicide when used as a single sign (Hansell et al., 1988) (Blends $M = 5.56$, $SD = 2.55$; Blends/R $M = 0.24$, $SD = 0.10$).

Mediation Section

The Mediation section uses a series of indicators to measure the extent to which the client is oriented toward making conventional, acceptable responses versus more unique ones. If either one of these directions is extreme and rigid, it suggests difficulties in adapting. This section includes simple frequencies for the total number of Populars and negative white space responses (S-; see previous interpretation for Populars and white space responses), along with the five percentages described next.

1. Conventional Form (X+%). X+% includes the form quality of all the responses in a protocol and, as such, tends to be less subject to distortions than F+% (see item 2, next). The X+% is essentially an indicator of the degree to which a person perceives things in a conventional, realistic manner. It is calculated as the sum of the number of

responses with a form quality of o or better (o or +) divided by the total number of responses.

$$\frac{\text{Sum of responses that have an FQ coding of o or +}}{R}$$

Most normal adults have an X+% of at least 68% ($SD = 11\%$; Exner & Erdberg, 2005). Normal children have a comparable mean, ranging between .67 and .78 (Exner & Weiner, 1995). An extremely high percentage ($> 90\%$) means that persons perceive their world in an overly conventional manner, to the extent that they might sacrifice their individuality. They are likely to be hypernormal, inflexible, rigid, and overly conventional (Exner, 1993, 2003). This is further supported by, and is consistent with, an elevated number of Populars. In contrast, lowerings in X+% ($< 70\%$) suggest persons who perceive their world in an unusual manner. This might be simply because they are highly committed to their individuality or, particularly if X+% is unusually low, it might suggest serious psychopathology. For example, patients with schizophrenia have a mean X+% of only 40% (Exner, 1993). There is very good research support for interpretations based on X+% (Hiller et al., 1999; Mihura et al., 2013) ($M = 0.68$, $SD = 0.11$; low $< .55$).

2. Conventional Pure Form (F+%). F+% assesses the same dimension as X+% but is limited to a narrower number of responses because it involves only pure F responses rather than other scoring categories (C', Y, T, and V) that might have been combined with F. It is calculated as:

$$\frac{\text{Sum FQ + and o}}{R}$$

Interpretation is similar to the interpretation of X+% but should be done more cautiously. It reflects a person's respect for the conventional aspects of reality and perceptual clarity. The Exner (1993) norms indicate that adults with schizophrenia have an F+% of only 42%, in contrast to the average of 71% among normals. In general, a low F+% might suggest limited intellectual endowment (S. J. Beck, 1961), organic impairment (Reitan, 1955b), or schizophrenia (S. J. Beck, 1968; T. Kahn & Giffen, 1960). There is relatively good research support for cautious interpretations of this percentage (Mihura et al., 2013).

3. Distorted Form (X-%). In contrast to X+% (and F+%), X-% is a direct index of the degree to which a person has distorted perceptions of reality. It is calculated as a percentage of the total number of responses that have a form quality of:

$$\frac{\text{Sum FQ-}}{R}$$

The higher the X-%, the more likely that the person has a significant level of impairment relating to distorted perceptions (Mihura et al., 2013). For example, moderately high percentages ($X\% = 20\%$) are found in depression, and percentages of 37% are characteristic in schizophrenia (Exner, 1993). Any percentage above 20% suggests that the person will have difficulty, because he or she has poor ties with reality and difficulty developing accurate abstractions. The research support for this variable is very strong (Mihura et al., 2013) ($M = 0.11$, $SD = 0.07$; high $> .20$, very high $> .30$).

4. White Space Distortion (S-%). Sometimes X+% and F+% can be low; it might then be assumed that this score is a result of a high number of form minus responses. This assumption might then result in incorrect interpretations. One way of checking for this difficulty is to note the percentage of minus responses for the white space (S-). Instead of suggesting the sort of distortions suggestive of schizophrenia (see interpretations for F+% and X+%), a low S-% might be caused by strong negativism or anger (Exner, 1993, 2003), but research supporting this is practically absent (Mihura et al., 2013).

5. Unusual Form (Xu%). Xu% also provides a check for potentially incorrect interpretations derived from low X+% or F+% scores. There might be cases in which X+% and F+% are low primarily as a result of a large proportion of unusual form (Fu) responses. The Xu% is calculated as a ratio of number of responses with form quality u compared to the total number of responses. Fu responses are unusual, but they still do not violate reality in the way that minus responses do, and thus they do not reflect severe pathology. In fact, a few Fu responses in a protocol can be a healthy sign that the person is capable of seeing his or her world in a novel manner; it may represent some aspect of creativity. However, an overabundance of Fu responses suggests the person is highly committed to an unconventional orientation (Exner, 1993, 2003). Unless the environment is highly tolerant of such an orientation, he or she is likely to have numerous conflicts and confrontations. The research supports this variable (Mihura et al., 2013) ($M = 0.20$, $SD = 0.09$; high $> .20$).

Processing Section

In addition to understanding clients' ideation and mediation, it is also important to assess the quality and efficiency by which they process information. Relevant frequency data are the overall amount of Organizational Activity (Zf; see interpretation under Organizational Activity), Perseveration (PER), Developmental Quality + (DQ+), Developmental Quality v (DQv), and the next three ratios.

1. Economy Index (W:D:Dd). The W:D:Dd ratio compares the degree to which an individual attempts to create a more challenging response that requires a high degree of organization and motivation (W), rather than choosing a less demanding and easily perceived area (D or Dd). It also relates a focus on the big picture to the focus on obvious or idiosyncratic details. It is simply a ratio of Whole responses (on the left side of the ratio) to D responses (in the middle) to Dd responses (on the right side). Normals and outpatients usually have a W:D ratio of 1:1.2 or even 1:1.8 (Exner, 1993). If a person includes a relatively large number of D responses, it suggests that he or she takes the least challenging way out of a conflict situation. It could be assumed that his or her characteristic way of dealing with ambiguity is to withdraw from it and focus on the obvious. If W is predominant, the person is perhaps overdriven in his or her attempts to organize perceptions. If, with a high W, both the W and the D responses are of poor quality, it suggests that a person is withdrawn and unrealistically striving for perfection (Exner, 1974). However, when W and D responses are both of good quality, they more likely represent the successful intellectual efforts of a creative person (Exner, 1974). Although this is an intuitively appealing set of interpretations, there is only minimal research supporting the Economy Index (Mihura et al., 2013).

2. Aspirational Ratio (W:M). The W:M ratio is a rough formula that, at the present time, is somewhat lacking in research (Mihura et al., 2013). As a result, interpretations should be treated with skepticism. It is a simple ratio of whole responses (placed on the left side of the ratio) to the total number of human movement responses (placed on the right side). It can be generally understood by reconsidering that the W response is an indicator of the degree to which subjects aspire to effectively organize and conceptualize their environments. It is an effort to encompass and include a number of different details in one coherent response. However, determining whether clients have the resources to actually accomplish an effective organization depends also on M. Although M represents the degree of investment individuals have in their fantasy lives, it also suggests how effectively they can bridge their inner resources with external reality and perform abstract thinking. Thus, the W:M ratio provides a rough comparison between a person's aspiration level, as represented by W, and his or her actual capability, as represented by M (Exner, 1993, 2003). Because introversives have higher M values than extratensives and ambitents, the relative value of EB needs to be taken into account in designating high or low W:M ratios. A high aspirational level is indicated if the W side of the ratio is greater than these values: introversives, 1.5:1; ambitents, 2.2:1; extratensives, 3:1 (Exner, 1993). However, scores with extremely high W components are common in children, which is consistent with the observation that children often underestimate the actual effort required to accomplish a goal (Exner, 1993, 1995). Ratios where the right side (M) is clearly lower than the left (0.5:1 for introversives and 1:1 for extratensives and ambitents) suggest that these persons are extremely cautious and conservative in defining achievable goals (Exner, 1993, 2003). Their motivation to achieve might be low, which would involve their being cautious (not wishing to fail), conservative in defining their objectives, and economical in their expenditure of energy. Again, there is little research on this variable currently (Mihura et al., 2013).

3. Processing Efficiency (Zd). Although the frequency of Organizational Activity (Zf), along with the Economy Index (W:D:Dd) and the Aspirational Ratio (W:M), provides information on the motivation and effort that persons place into their perceptions, these indicators do not provide information related to quality or accuracy. In contrast, the Processing Efficiency (Zd) score provides an index not only of effort but also of ease and accuracy of processing. It is calculated as a difference score, subtracting Zest (summing the total number of times an Organizational Activity response occurred in a protocol) from ZSum (which incorporates the Z score weightings). Individuals scoring high on Zd are considered to have an overincorporative style; they invest more effort and are more accurate in their perceptions and conclusions. This seems to be an enduring traitlike feature. In contrast, low scorers have an underincorporative style, which means that they process information in a more haphazard style, often neglecting relevant bits of information. This characteristic seems more amenable to change, as indicated by moves to a more overincorporative style following psychotherapy (Exner, 1978). A review of research on Zd (Exner, 1993) indicated that, consistent with theory, overincorporators (high Zd) have more extensive eye-scanning, make fewer errors on games, and are less likely to make guesses related to requests for factual information. In contrast, underincorporators (low Zd) make fewer eye movements while scanning, are more likely to make errors on games, and are more likely to make guesses related to factual information. Among children, low Zd scores occur among those diagnosed as

hyperactive. However, the preponderance of research has not supported the use of Zd (Mihura et al., 2013), so practitioners should consider not interpreting this variable.

High Zd (> +3)

- Possibly obsessive or perfectionistic; can also efficiently and accurately process information.
- Exert more effort in information processing.
- Will take care with their perceptions and continually check for accuracy.
- Confident in their abilities.

Low Zd (< -3)

- Likely to be haphazard, might make impulsive decisions without fully taking into account all relevant aspects of a situation.
- Will invest minimal effort into actively working with their perceptions.
- Are typically uneasy with their information-processing ability, may question their efficiency at perceiving, integrating, and responding to information.

Interpersonal Section

Although the Rorschach does not obtain information regarding a person's actual environment or the other persons in that environment, it does provide information related to needs, attitudes, behavioral response sets, and coping styles, all of which are relevant to interpersonal relationships. The Interpersonal section lists several measures relevant to these domains. The person's degree of cooperation with others can be noted through his or her number of Cooperative Movements (COP). Similarly, the total number of Aggressive Movements (AG) provides an index of interpersonal aggression, and a high amount of Food contents suggests dependency (see interpretations under each of these categories). Additional useful indicators of interpersonal relations include sum of pure H, number of Perseverations (PER), ratio of Good to Poor Human Representations (GHR:PHR), sum T, and active:passive responses (see interpretations under each one of these categories). The next two formulas can also be useful in assessing the extent to which a person is interested in people as opposed to being isolated.

1. **Interpersonal Interest [H + (Hd) + Hd + (Hd)].** The Interpersonal Interest index merely adds up the total amount of human content (not including Human Experience; Hx). It represents the degree to which a person is interested in other people, but in general research does not support interpreting this variable with confidence (Mihura et al., 2013) ($M = 6.29$, $SD = 2.66$).
2. **Isolation Index (Isolate/R):** Exner (1986) pointed out that the five contents (Botany, Clouds, Geography, Landscape, and Nature) used to develop the Isolation Index are all "nonhuman, nonsocial, inanimate, and usually static objects" (p. 406). This index is calculated by summing the total number of each of these contents, with clouds and nature weighted times 2, then dividing by the total number of responses:

$$\frac{Bt + 2Cl + Ge + Ls + 2NA}{R}$$

If a high proportion of these contents (index score of .25 or greater) occurs in a person's protocol, it suggests the person may be withdrawn or alienated or may at least have some difficulties related to social isolation (Exner, 1993). This applies to children, adolescents, and adults (Exner, 1986, 1995). However, these interpretations should not necessarily take on a pathological bias. A high score might merely represent less interest in people rather than a negative rejection and alienation from them. Research does not support interpreting this variable, though (Mihura et al., 2013) ($M = 0.19$, $SD = 0.09$).

Self-Perception Section

The Self-Perception section includes information relevant to the relative assets and limitations of the clients as seen by the clients themselves. These entries are simply frequency tallies: Fr + rF, Form Dimension, sum of Morbid content, Anatomy/X ray responses, and sum V (see interpretations under relevant sections). The ratio of Pure H:(H) + Hd + (Hd) compares the amount of Pure Human responses with mythical/fictional and part human responses. Two of the human categories on the right side of the ratio relate to fictional/mythical descriptions. As such, they can be considered to represent the extent to which the individual bases his or her perceptions on real versus imaginary aspects of people. Adult and adolescent nonpatients usually give more Pure Human responses than (H) + Hd + (Hd) at a rate of approximately 3:2 (Exner, 1993, 1995, 2003). However, the means for the ratio are different for introverts (3:1) than for either extroverts or ambiverts (1.3:1). In contrast, people with schizophrenia see a much higher proportion of fictional/mythical and part-human responses (1.5:2; Exner, 1993). This low a level of Pure Human responses suggests that they are working from an unrealistic perception of themselves and others. The Self-Perception section also includes one quantitative formula:

1. **Egocentricity Index [3r + (2)/R].** The Egocentricity Index (EI) provides information related to whether the client has a sense of self-worth and further relates this to the extent that he or she is absorbed with self. It gives three times more weighting to reflection responses than to pair responses, adds these up, and divides by the total number of responses:

$$\frac{3(\text{Fr} + \text{rF}) + \text{Sum}(2)}{\text{R}}$$

However, research does not support interpreting this variable (Mihura et al., 2013), so practitioners should consider not using the below interpretations.

High EI (> .44)

- Overinflated sense of self-worth, which reflects underlying dissatisfaction.
- Moderate elevations (index level of .40 to .45) indicate self-focusing and self-concern associated with positive self-esteem.

Low EI (< .33)

- Negative sense of self-worth, conflicted self-image, possible mood fluctuations and dysfunctional behaviors.

Special Indices

In an attempt to increase the robustness and validity of various combinations of Rorschach measures, six Special Indices have been developed based on a composite of scores. For example, a number of different indicators of schizophrenia are found throughout the Rorschach. These include a high number of X-% or M-%, the presence of one or more Level 2 Fabulized Combinations (FAB2), and a high WSum6. These, along with several other indicators, were combined to form the Perceptual-Thinking Index (PTI). Some research has found that this index can discriminate schizophrenia better than any of the single scores (see Exner, 1991, 1993, 2003). A similar strategy was used for the other Special Indices. Collectively, they help to form a nucleus of indices to help with more specific types of diagnostic conditions. Exner (1993, 2003) and the commercially available scoring forms have included Constellation Worksheets for calculating whether the Special Indices are positive (also see Exner, 2003, Table 10.5, p. 156). These calculations are significantly more complicated than the others presented in this chapter, so details of scoring them are not presented here.

1. Perceptual Thinking Index (PTI). The PTI is a revision of the earlier Schizophrenia Index (SCZI). PTI has the advantage of more accurately identifying persons with thought disorders (Exner, 2003; Dao & Prevatt, 2006; S. R. Smith, Baity, Knowles, & Hilsenroth, 2002). As the name suggests, it is not designed to diagnose schizophrenia but more to assess the array of disorganized or unusual thought processes and perceptual experiences that may occur with schizophrenia or other forms of thought disorders. It should also be considered as rating a person on a continuum of thought disturbances rather than being designed to place a person in a certain category (diagnosis). This is a well-supported variable (Mihura et al., 2013). The recommended cutoff score for determining problematic thinking is = 3 (Dao & Prevatt, 2006).

2. Depression Index (DEPI). A DEPI value of 4 raises the possibility that the client is experiencing some depressive symptoms—fluctuations in moods, a sense of dissatisfaction, pessimism, and some mild vegetative symptoms (fatigue, insomnia, slowed thinking, anhedonia). Scores of 5, 6, or especially 7 are far more definitive and strengthen the likelihood of an affective disorder as reflected by an intensification of these symptoms (Exner, 1991, 1993; Exner & Erdberg, 2005). However, the specific diagnosis of a depressive disorder may not be warranted because depression is generic to a wide variety of disorders, particularly many of the personality disorders and schizophrenia. In addition, the term *depression* might be used to describe people who are emotionally distraught or are pessimistic, self-defeating, and lethargic, as well as those who feel a sense of futility when attempting to function competently in a complex society (Exner, 1993). Depressive symptoms and tendencies as measured by DEPI may, therefore, relate to both a wide number of types of people and a wide range of possible diagnoses. This index should be used with considerable caution because it has been identified by a number of authors as having questionable validity, especially with children and adolescents (Ball et al., 1991; Jorgensen et al., 2000; Meyer & Archer, 2001; Stredny & Ball, 2005). Overall, while the preponderance of research has indeed supported this index, the effect size is low (Mihura et al., 2013).

3. Coping Deficit Index (CDI). It has been conceptualized that clients with scores above 4 or 5 on the CDI are likely to have unsatisfying and somewhat meaningless

interpersonal relationships, largely because they find it difficult to effectively deal with everyday requirements (Exner, 1993). Their histories typically include social ineptness, poor success in interpersonal relationships, and times when they have felt overwhelmed by interpersonal demands. Effective moderate- to long-term psychotherapy was found to result in decreases in CDI (Exner & Sanglade, 1992; Weiner & Exner, 1991). Similar to DEPI, the overall research supports this index, but with a low effect size (Mihura et al., 2013), so interpretations should be made with caution.

4. Suicide Constellation (S-CON). The Suicide Constellation includes 11 variables that collectively are intended to detect persons at risk of attempting suicide. Retrospective studies indicate that, using a cutoff score of 8, 80% of suicide attempters were identified accurately (Exner, 1986, 1993). However, caution should be exercised in making final decisions. Some clients were incorrectly identified as not being suicidal and yet they later made attempts (false negative rate = 15%). Among depressed populations, a number of clients were incorrectly identified as being at risk of suicide when there was actually no or little risk (false positive rate among adults with depression = 10%; Exner, 1993). Despite this, there is generally excellent research support for this variable as a measure of risk (Mihura et al., 2013).

5. Hypervigilance Index (HVI). Originally, a series of indicators was isolated from patient protocols that seemed to differentiate paranoid-type patients (paranoid schizophrenia) from other patient groups. This series of indicators was partially successful in that persons with paranoid schizophrenia and paranoid personalities were identified correctly (88% and 90%, respectively; Exner, 1993). On further investigation, it was found that HVI related more to the hypervigilant aspect of the paranoid style rather than to paranoia itself. Thus, persons with positive indicators on HVI are likely to place a large amount of effort into maintaining a high state of preparedness. Motivating this is a sense that they mistrust their environment and experience a chronic sense of vulnerability (Exner, 1993). Before initiating behaviors, they carefully think through why and how they should express them. They are likely to be quite guarded regarding closeness in relationships and initially respond to efforts at closeness with apprehension. As a result, they allow themselves to be close with others only if they feel in control. They are generally quite concerned with issues not only of emotional closeness, but also of personal space in general (Exner, 1993). It is important to note from the Mihura et al. (2013) meta-analysis that not only does the research generally not support interpreting this variable, but a significant but low effect size was found in the opposite direction from what was expected, suggesting that interpretation of this index is particularly problematic.

6. Obsessive Style Index (OBS). The Obsessive Style Index was developed by examining the records of clients who had been formally diagnosed as having obsessive-compulsive disorder to determine which Rorschach characteristics could distinguish them from other groups. Five characteristics were isolated, and, using the designated criteria, they correctly identified 69% of those with obsessive-compulsive disorder (Exner, 1993). If the OBS is positive (score of 3 or more), it suggests persons who are perfectionistic, indecisive, and preoccupied with details and who experience difficulty expressing emotion. They are likely to be cautious, conservative, conforming, and conventional (check for high Populars). They process information extremely methodically and, when using the Zd Index definition, are likely to be overincorporators

(see interpretation for Zd). However, a positive index does not necessarily indicate psychopathology; rather, it shows a style of approaching the world and processing information. If this style is overly rigid, it can become dysfunctional, particularly when the person is under pressure or is required to achieve goals within a limited time (Exner, 1993). However, the preponderance of research does not support interpreting the OBS (Mihura et al., 2013).

R-PAS: ADMINISTRATION

As with the Comprehensive System, examiners should standardize their administration procedures as much as possible, as administration changes can influence client responses. Examiners should minimize the variations in their administration procedures as much as possible. As with the Comprehensive System, the examiner should sit side by side with the respondent if possible, or corner to corner at least. This minimizes the possible influence of subtle cues the examiner may convey with his or her face. Once ready, the examiner should utilize the following sequence of steps, derived from Meyer et al. (2011).

Step 1: Establishing Rapport

Establishing a positive working alliance with a client is extremely important with a task that is as ambiguous, opaque, and potentially frustrating as the Rorschach. The key is to build motivation to fully engage with the task, and the relationship between the assessor and client can be a key motivator or inhibitor. The R-PAS manual (Meyer et al., 2011) suggests asking informal and neutral questions, administering easy drawing or other performance-based measures (e.g., projective drawings or even the Bender–2), or doing a brief, non-emotionally arousing clinical interview, for demographic or biographical information. Small talk can be a simple but effective way to build rapport, as can the use of humor. It is important to note, though, that whatever happens just before the administration of the Rorschach may affect the client's responses. It is best to avoid emotionally charged topics when talking just before the Rorschach and before the administration of other challenging tests and measures. Administering a self-report measure that asks questions about suicidality and self-harm may trigger specific feelings and/or imagery for certain individuals, which may carry over into the Rorschach responses. There is no way to control everything that may leave residual images in the client's mind, but examiners can at least minimize the impact they have on this by being aware of what happens just before administration. Although there are many key aspects to building a positive relationship, alliance, and rapport (see Norcross, 2011), expressing genuine concern and curiosity about the client, even during small talk, can build the basis of a relationship that will in turn encourage effortful engagement on the part of the client.

Step 2: Giving the Testing Instructions

The major underlying principle in introducing the test and its instructions to the client is to let the client lead the process as much as possible. The R-PAS manual

(Meyer et al., 2011) instructs evaluators to first introduce the test and ask if the client has ever heard of it, seen it, or taken it before. If the client has, the examiner has the task of briefly exploring what the client knows (or thinks) about the Rorschach and correcting any misconceptions about the task. It is important to note that this discussion should not be about the benefits of the test, what it measures, or other things that have nothing to do with the actual task of the test, which is for the client to say what the inkblots look like to them. Regardless of whether or not they have prior knowledge of the test, the examiner must establish that they understand the task itself. As with the Comprehensive System, the task of the client is to look at each inkblot and answer this question: “What might this be?”

A change from the Comprehensive System administration procedure is the addition of a response expectation of two or three responses. The R-PAS instructions direct the examiner to say: “Try to give two responses ... or maybe three, to each card. That is, for each card try to see two different things; possibly three” (Meyer et al., 2011, p. 8). This more directive expectation accomplishes two goals. First, it proactively aims to restrict the range of number of responses to an optimal level, between 18 and 27 in total. Second, it serves as a warning for the prompt and pull procedures that may occur during the response phase (see “Step 3: The Response Phase”).

As with the Comprehensive System, any questions that arise during the R-PAS should be handled directly, succinctly, honestly, and as nondirectively as possible. Postponing answering more in-depth questions about the utility or function of the test is acceptable (e.g., “Let’s talk about that when we’ve finished the testing”). Direct, honest responses that work for multiple purposes include: “Different people see different things” and “That’s up to you.” It is important not to be misleading or cryptic, as this could affect rapport. Examiners should understand that the process of taking the Rorschach can be frustrating for many people, and even an empathic tone of voice when not directly giving a satisfying answer to a question can help.

Step 3: The Response Phase

Throughout the testing procedure, the basic conditions of step 2 should be adhered to as closely as possible. However, specific situations often arise as examinees are free-associating to the Rorschach designs. If a respondent requests specifics on how to respond or asks the examiner for encouragement or approval, examiners should consistently reply that the subject can respond however he or she likes.

The first goal of this phase is for the examiner to structure the test, by handing the cards, asking the “What might this be?” prompt, administering the prompts and pulls procedures, as necessary (discussed just below). The second goal is for the examiner to record the responses and any notable nonverbal responses. As with the Comprehensive System, verbal responses should be recorded verbatim. To simplify this process, most clinicians develop a series of abbreviations. A set of abbreviations used throughout all the Rorschach systems consists of the symbols (\vee , $<$, \wedge , $>$) in which the peak indicates the angle of the card. It is also important to note any odd or unusual responses to the cards, such as an apparent increase in anxiety, wandering of attention, or acting out on any of the percepts. However, recording the verbatim verbal response should always take precedence and should never be compromised in order to record other behaviors.

The prompting and pulling procedures that have been added to the R-PAS are also written about as the R-Optimized administration procedure. They were created to dramatically decrease the odds that invalidly short (15 or fewer responses) protocols and extremely lengthy protocols (which for the most part do not add any new or useful information after a certain point) emerge. Prompts are used to increase the number of responses. The first time a respondent offers only one response to a card, the examiner should prompt the respondent, reminding him or her that two or maybe three responses were requested. A prompt should be given to each card that receives only one response, though the wording can change slightly each time and include phrases like “Take your time” and “I wonder if you can see something else there too.” Prompting should only occur once per card, even if the respondent does not offer a second response.

The pulling procedure is meant to help limit lengthy protocols. If a client gives four responses to a single card, the examiner should ask him or her to return the card and move on to the next one, reminding him or her that two or maybe three responses is enough. This should happen every time the respondent gives four responses to a single card, though after repeated pulls, the examiner can narrow down the feedback to let the respondent know that two responses is enough. It is important to note that examiners should never sacrifice rapport or act in a way that is likely to upset a client during the R-PAS. As such, if an examiner invokes the pull procedure but the client insists on providing more responses to that card, the examiner should allow this (even though the extra responses will not be clarified in the next phase or coded). It is hoped that the range of number of responses during the R-Optimized administration is between 26 and 40 or, even better, between 18 and 27.

Step 4: The Clarification Phase

The Clarification Phase begins after all 10 cards have been administered. Its purpose is to collect the additional information required for an accurate coding of the responses. It is intended to clarify the responses that have already been given, not to obtain new responses. The information needed from the clarification phase should ensure that the examiner can code the response. Clarification questions and follow-up are required when additional codes are implied (usually with key words) or if required codes are not addressed (such as when a respondent does not articulate a location). The R-PAS manual (Meyer et al., 2011) recommended that the instructions for the inquiry closely approximate these ones:

Now we are going to start the final step. While looking at the cards I want to review your responses with you to clarify what it is that you saw and how you saw it. So we will look at the cards one by one. I will read your responses back to you and I want to know where on the card you were looking and what about the inkblot made it look like that to you. Does that make sense? (p. 16)

Following closely the general theme of the overall administration, the inquiry should not influence the examinee's responses. Thus, any questions should be as nondirective as possible. The examiner should begin by merely repeating what the respondent has said and then waiting. Usually the respondent begins to clarify his or her response.

If this information is insufficient to clarify how to code the response, the examiner might become slightly more directive by asking, “What about it made it look like [a percept]?” A good working knowledge of the coding system of the R-PAS is required in order to accomplish the clarification adequately. The examiner should *not* ask, “Is it mainly the shape?” or “How important was the color?” These questions are far too directive and are worded in a way that can exert influence on the respondent’s descriptions of his or her responses. The examiner should consistently avoid leading the client or indicating how he or she should respond. Particular skill is required when clarifying a determinant that has been unclearly articulated but merely implied.

The outcome of a well-conducted Clarification Phase is the collection of information sufficient to decide on coding for each response. However, the examiner should not sacrifice rapport during this phase in order to get the information. Examiners must make judgment calls throughout this phase for how best to elicit the information needed from clients. This may mean trying different leading questions during the inquiry until one seems to be effective, then using that one (or slight variations on it) for the rest of the administration. It can be difficult for examiners to find ways to get the necessary information in the Clarification Phase in a supportive way that minimizes frustration as much as possible.

R-PAS: CODING

The next step following administration is to code the different categories for each response. Many of the codes are similar, if not the same, as those used in the Comprehensive System. Some have been altered, though, and some are new. The subsections that follow merely list, outline, and define the coding categories. To achieve accurate coding (and scoring), practitioners should consult the *Rorschach Performance Assessment System Administration, Coding, Interpretation, and Technical Manual* (Meyer et al., 2011), which includes specific coding criteria, tables, charts, and diagrams. The inclusion of specific scoring criteria is beyond the scope of this chapter. The focus here is on providing a key to interpretation that is concise, accountable, and clearly organized. The definitions and the accompanying tables serve to outline and briefly define the primary R-PAS factors.

Location and Space

The *location* of the responses refers to the area of the inkblot that is used (Table 11.9). This can vary from the use of the entire blot (whole response) to the use of an unusual detail (Dd). Unusual details are defined as location responses made by less than 5% of normative sample subjects. The use of white space is coded differently on the R-PAS from the Comprehensive System. Instead of white space (S) being coded with location, it has a separate coding section on the record form. This section breaks the use of white space into *white space reversal* (SR), which is used when the respondent uses the white space as the primary figure in the response, and *white space integration* (SI), which is used when the respondent integrates the white space on the card into a larger response that uses the inked portion of the card as the primary figure.

Table 11.9 Location and Space Codes

Code	Location Name	Description
W	Whole	Response uses the entire inkblot
D	Common Detail Area	Response uses one or more frequently used detail areas
Dd	Unusual Detail	Response uses one or more rarely used detail areas
<i>White Space</i>		
SR	White Space Reversal	Non-inked or background area on card is a focal percept such that the traditional figure and ground become reversed
SI	White Space Integration	Non-inked or background area on card is integrated with inked areas

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

Content

The coding of content uses some of the same codes as the Comprehensive System, but it has pared down the number of possible contents based on empirical support for each (and for the calculations that use them). Table 11.10 provides a listing of the content categories included in the R-PAS, with the code symbol/abbreviation and description for each category. When more than one content category occurs in the same response, all contents should be coded. The R-PAS includes a miscellaneous (NC) code for all the contents that do not otherwise fit into a category.

Object Qualities

Three separate qualities have been extracted from the Comprehensive System and simplified in their coding for the R-PAS (Table 11.11). First, the Synthesis (Sy) code is applied when two or more separate objects or contents in the response are articulated as being in some sort of significant relationship with each other. Any meaningful interaction or relationship between more than one object is coded with Sy. Vagueness (Vg) applies to responses in which all of the content lacks what is known in the Comprehensive System as form demand. That is, the objects do not have a fixed, specific shape to them. For example, a tree has a basic, fixed shape (even though that shape may be variable based on the type of tree), so it is not considered vague. However, a cloud does not have a fixed, specific shape (nor do fire, water, smoke, etc.), and so would be coded with a Vg. If any objects in the response do have a specific, consistent shape, Vg is not coded. For example, a response that included water, fire, and a spider would not be coded Vg, as a spider has a distinct, consistent shape to it.

The Pair (2) code applies when there are two separate, identical objects that are seen on either side of the midline of the blot. That is, the two identical objects must be based on the symmetry of the blot. They cannot be two identical parts of a single object (e.g., two wings on a bat).

Table 11.10 Content Codes

Code	Description
H	Whole human, including realistically described religious or historical figures; e.g., person, baby or fetus, Buddha. Also code figures described as humans but with non-human parts; e.g., a man with wings.
(H)	Imaginary, fictional, quasi-, or supernatural whole human; e.g., ghost, giant, human-like monster, demon, fairy, God, clown.
Hd	Human detail, for an incomplete human form; e.g., face, head, leg, mustache (but not those percepts better captured as Anatomy, as described below).
(Hd)	Imaginary or fictional human detail; e.g., face of the devil, angel's wing, human-like mask, carved pumpkin face/jack-o-lantern.
A	Whole animal; e.g., butterfly, elephant, insect, pterodactyl, amoeba, a bat with hands.
(A)	Imaginary, fictional, or cartoon whole animal; e.g., unicorn, King Kong, Nemo, teddy bear, dragon, animal-like monsters.
Ad	Animal detail, for an incomplete animal form; e.g., pelt, head, wing, antler (but not those percepts better captured as Anatomy, as described below).
(Ad)	Imaginary or fictional animal detail; e.g., wings of a dragon, animal mask.
An	Anatomy, for internal body parts and structures that are not visible from the outside; e.g., pelvis bone, intestines, brain cross-section, lungs, cells, cross section of an organ. Also for perceptions of anatomy from medical imaging devices, including MRI, PET scan, X-ray, or ultrasound technology. Internal human or animal body parts that are coded An are not also coded Hd or Ad, unless external body parts are also included.
Art	Art, for objects of art, e.g., paintings; or for objects that are, or are described as being, decorative or ornamental, e.g., crest, jewelry, ceremonial feathers, a fancy and delicate table.
Ay	Anthropology, for references to a specific historical or cultural context; e.g., Indian arrowhead, Napoleon's hat, Mongolian yurt, totem pole.
Bl	Blood.
Cg	Clothing; e.g., bowtie, dress, boots, hat.
Ex	Explosion, including bomb blast, volcanic eruption, and fireworks.
Fi	Fire, including flames, embers, or smoke.
Sx	Sexual organs, activity, or clothing; e.g., penis, vagina, see-through dress.
NC	Objects and contents that are not classified in other categories; e.g., cloud landscape, and lamp; including abstractions like depression and sensory experiences.

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

Form Quality

Like in the Comprehensive System, the R-PAS includes a measure of how well a content “fits” a particular part of the inkblot, based on how frequently similar content was observed by the normative sample (Table 11.12). When all of the content in the

Table 11.11 Object Qualities

Code	Name	Description
Sy	Synthesis	Distinct and separate objects in a relationship
Vg	Vagueness	Objects with vague or indistinct outline or boundary
2	Pair	Identical objects based on the symmetry of the blot

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

Table 11.12 Form Quality

Code	Name	Description
o	Ordinary	Form fit that is relatively frequent and accurate
u	Unusual	Form fit that is of intermediate frequency or accuracy or both
–	Minus	Form fit that is infrequent and inaccurate
n	None	Response does not contain an object with definite form or outline

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

blot is vague (formless), and form is not used at all as a determinant to explain why and how the blot or blot area looks like that content, a code of FQn (none) is applied. This should be the first decision for each response, whether there is any form quality or not. If there is any form used in the response as a determinant or if there is a specific shape to any of the content (i.e., Vg is not coded), then a comparison must be done between the response content and the R-PAS tables (chapter 6 in the R-PAS manual; Meyer et al., 2011) for content responses for the same location. Ordinary form quality responses (FQo) are contents that occur quite frequently for the specific location in the normative sample and as such are quite conventional, logical uses of the inkblot contours and qualities. Unusual form quality responses (FQu) occur less frequently but do not represent perceptual distortions of the inkblot contours and qualities. Minus responses (FQ–), however, are rare and often quite difficult for others to see as the respondent did. They may be distortions of the blot stimulus or arbitrary use of some of the blot qualities, ignoring others.

Popular Responses

The R-PAS Popular (P) code uses the same definition and list as the Comprehensive System, referring to the presence of frequently perceived contents in responses. Exner (2003) used, as the cutoff for inclusion as a Popular, an occurrence of at least once in every three protocols from nonpsychiatric populations. This list of Popular responses is detailed in Table 11.13.

Table 11.13 Popular Responses

Card	Location	Description of Popular
I	W	Bat or butterfly. If head is mentioned, it must be at top of card (e.g., by D1)
II	D1	Bear, dog, elephant, or lamb; usually head or upper body
III	D9	Whole human figure or representation of a human figure (e.g., doll)
IV	W or D7	Whole human or human-like figure [i.e., coded as H or (H)]
V	W	Bat or butterfly. If head is mentioned, it must be at top of card (D6 area)
VI	W or D1	Animal skin, hide, rug, or pelt
VII	D9	Human head or face (often the response contains other features elsewhere)
VIII	D1	Whole animal, with head near D4. The type is often unspecified, though bear is common, as are various times of canines, felines, or small mammals.
IX	D3	Human or human-like figure
X	D1	Crab or spider

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

Determinants

As in the Comprehensive System, *determinants* refer to the characteristics either of the blots or attributed to the blots to which the examinee responds, such as its shape, color, movement, or texture (Table 11.14). They are the reason an examinee gives that the blot looks like whatever he or she responded as having seen. The coding system for determinants in the R-PAS is significantly simpler than that in the Comprehensive System, as chromatic color is the only code for which the coder must establish whether color or form is primary. The other codes (e.g., V, T, Y, FD) are simply applied if they are present. Additionally, the form (F) code is only applied when form is the *only* determinant; it is never part of a blend in the R-PAS.

As with the Comprehensive System, all movement responses should be coded for the extent to which the movement is active versus passive. Active movement would include movements such as “fleeing” or “lifting,” whereas more passive movements might include “meditating” or “anchored,” with “talking” being the cutoff on the continuum between passive and active (talking is coded as passive, so decisions should be made whether the movement is more passive than talking, which would get a *p*, or whether it is more active than talking, which would get an *a* code). Whether a movement is active or passive is designated with either an *a* (for active) or a *p* (for passive) superscript. The *a* and *p* designations are later scored and used for interpretation in the quantitative summaries (see the “Ideation Section” topic in the section titled “Structural Summary”).

Table 11.14 Determinants

Code	Name	Description
<i>Movement</i>		
M	Human Movement	Human activity, experience, sensation, and emotion
FM	Animal Movement	Animal activity, and animal experience; animals in movement
m	Inanimate Movement	Mechanical or inorganic movement, including natural forces
<i>Type of Movement</i>		
a	Active	The amount of effort or force incorporated in a movement
p	Passive	
<i>Color</i>		
FC	Form Color	Color contributes to a response object but form is dominant
CF	Color Form	Color is dominant in a response object but form contributes
C	Color (no form)	Color determines a response object without form playing a part
C'	<i>Achromatic Color</i>	Black, grey, or white color of the ink contributes to the response
<i>Shading</i>		
T	Texture	Light and dark ink gradations contribute to a response and contribute to a tactile quality
V	Vista	... and contribute to a perception of depth or dimensionality
Y	Diffuse Shading	... but do not contribute to a tactile impression or a sense of depth
FD	<i>Form Dimension</i>	Blot outlines generate a perception of depth or dimensionality
r	<i>Reflection</i>	An object plus its mirror image or reflection across the card's vertical midline
F	<i>Form</i>	Responses in which form is the only determinant

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

As in the Comprehensive System, the R-PAS allows for the complexity of responses to be coded with *blends*, which occur when more than one determinant is used in a single response (with the exception of form, which is not coded in a blend). All determinants used should be included in the coding of a response.

Cognitive Codes

The R-PAS retained from the Comprehensive System the first six Special Score categories that were developed to evaluate problematic thought processes and confusion. These have been renamed Cognitive Codes in the R-PAS, but they are coded

Table 11.15 Cognitive Codes

Type	Code	Name
Language & Reasoning	DV1	Deviant Verbalization Level 1
	DV2	Deviant Verbalization Level 2
	DR1	Deviant Response Level 1
	DR2	Deviant Response Level 2
	PEC	Peculiar Logic
Perceptual	INC1	Incongruous Combination Level 1
	INC2	Incongruous Combination Level 2
	FAB1	Fabulized Combination Level 1
	FAB2	Fabulized Combination Level 2
	CON	Contamination

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

(and scored with weights) similarly. These, along with their definitions, are listed in Table 11.15. As in the Comprehensive System, four of the Special Scores (DV, DR, INCOM, and FABCOM) are coded at a Level 1 if the response is only mildly atypical and at a Level 2 if the response is more bizarre or unusual, indicating more likely cognitive disruption.

Deviant verbalizations (DV) refer to linguistic abnormalities, such as made-up words (neologisms) and misused words (malapropisms). The relative comprehensibility distinguishes a DV1 from a DV2. For example, a respondent misusing a word that closely resembles or sounds like the intended word may be coded a DV1. Similarly, a logically made-up word, such as “unperturbable,” would be a DV1. More bizarre made-up words or confused word use, resulting in difficulty understanding the meaning, would be coded as DV2. Deviant responses (DR) refer to phrases or language that are off task, not helping explain what the individual sees in the blot or why. This may take the form of rambling or odd and unnecessary details. The level (DR1 versus DR2) depends on the degree of confusion or how off task the statements are. Peculiar logic (PEC), renamed from Inappropriate Logic (ALOG) in the Comprehensive System, relates to spontaneous strained or unusual logic used to justify a response. This often comes in the form of a “because” statement, justifying that it is what the client said it is *because* of some peculiar logic.

Incongruous Combinations (INC) are coded when an object has an implausible component or aspect to it, such as a body part that would not actually be found on a body or a characteristic that an object would not actually have. The degree of bizarreness or lack of logic for the combination distinguishes an INC1 from an INC2. Fabulized Combinations (FAB) are coded when two or more objects have an implausible relationship or engagement. Again, the degree of bizarreness or lack of logic for the relationship distinguishes a relatively understandable response (FAB1) from a truly bizarre one (FAB2). Contaminations (CON) are extremely rare and are coded when

two completely separate and distinct images are perceived and superimposed onto each other, merged together in the same blot location.

Thematic Codes

The R-PAS also includes nine Thematic Code categories (Table 11.16) that were developed to take into account characteristics of the response not covered in any of the other codes, some of which are used in the Comprehensive System and some of which are not. Inclusion of symbolic and representational material in a response (ABS), use of personal knowledge to justify a response (PER), cooperative and aggressive movement (COP, AGM), and the presence of morbid, damaged, broken, distressed, or dysphoric representations (MOR) are all similar to Comprehensive System codes. Additionally, Good and Poor Human Representations (GHR and PHR) are coded automatically by the R-PAS online system, using the same complex algorithm as in the Comprehensive System. Mutuality of Autonomy, either Health (MAH) or Pathology (MAP), represents qualities of relationships between two or more objects in the blot or between an object in the blot and something outside of it. MAH represents relationships that

Table 11.16 Thematic Codes

Code	Name	Description
ABS	Abstract Representation	Concrete blot features are representational and symbolize an abstract, higher-order construct or concept.
PER	Personal Knowledge Justification	Personal knowledge or experience is used to justify a response.
COP	Cooperative Movement	Cooperative, positive, or pleasant interactions are occurring between two objects.
MAH	Mutuality of Autonomy-Health	Two objects are mutually and autonomously engaged in a reciprocally interactive activity.
MAP	Mutuality of Autonomy-Pathology	An agent or object intentionally compromises the autonomy or integrity of another object or is destructive to it.
AGM	Aggressive Movement	Aggressive or hostile activity, intent, or ideation is occurring.
AGC	Aggressive Content	Response content involves an aggressive, dangerous, harmful, injurious, malevolent, or predatory element.
MOR	Morbid Content	Objects are damaged or states of distress or dysphoria are attributed to them.
ODL	Oral Dependency Language	Response Phase verbalizations linked to oral activity and content or interpersonal passivity and dependence.

Source: Reproduced from the *Rorschach Performance Assessment System: Administration, Coding, Interpretation, and Technical Manual*, © 2011 by Rorschach Performance Assessment System LLC. All rights reserved. Used by permission of Rorschach Performance Assessment System LLC.

are reciprocally interactive, and the objects that are positively interacting must both be in the blot itself. MAP represents a relationship in which one object (which can be in the blot or not) is negatively impacting or has negatively impacted the autonomy of another object (which can be in the blot or not). That is, one object has harmed, aggressed against, dominated, controlled, or intends to do so to another object. The aggressive content (AGC) code applies when an object seen in the blot is generally perceived to be harmful, dangerous, or malevolent.

Oral Dependent Language (ODL) is the only code in the R-PAS that is applied only using the responses in the Response Phase (ignoring the Clarification Phase). The ODL code is applied for two different types of descriptions: (1) any language that is associated with oral imagery, such as talking, eating, food, mouth, or other mouth or eating involved language; and (2) any language that is associated with dependency, such as begging, relying, or kneeling before someone or something.

R-PAS: SCORING THE STRUCTURAL SUMMARY

After the examinee's responses have been coded according to the categories just presented, codes can be entered into the Internet-based scoring program at www.r-pas.org. Like the Comprehensive System, the R-PAS produces quantitative formulas comprised of various ratios, percentages, and derivations. These formulas reflect the proportions of, and comparisons among, various Rorschach factors. After the quantitative formulas have been calculated, they become the primary focus on which Rorschach interpretations are made. The calculations are presented by degree of empirical support and thus interpretive confidence (what the R-PAS calls *Page 1* variables, which have the most significant empirical support, versus *Page 2* variables, which have promising empirical support) and then by four domains: Engagement and Cognitive Processing, Perception and Thinking Problems, Stress and Distress, and Self and Other Representation. These domains are in addition to some preliminary information on administration behaviors and observations. Each score is presented by the R-PAS scoring program both numerically and graphically. Standard and percentile scores are presented, and a continuum representing scores from low to high is presented for each variable. In addition to being placed on the line graph, each point is color- and shape-coded, such that green markers with no lines through them represent scores that are within normal limits (generally between a standard score of 90 and 100), yellow markers with one line through them are slightly outside of this normal range, red markers with two lines through them are significantly outside of the normal range (70–80 or 120–130), and black markers completely filled in are outside of 2 standard deviations beyond the mean (below 70 or above 130). Markers that are grayed out are not applicable.

The R-PAS has the ability to provide two different sets of scores for each of the variables (ratios/formulas). The raw score option is often the more easily interpreted, as it works like most other standardized tests, comparing the respondent's scores to the normative sample and resulting in a standard score and percentile rank. The standard scores on the R-PAS work like IQ scores, with a mean of 100 and a standard deviation of 15, making interpretation relatively straightforward. The second option

for calculation and display is adjusted based on the Complexity variable. The Complexity variable, which is the first factor calculated by the R-PAS, includes numerical point accrual based on how complex responses are in location (more complex responses include synthesis, white space), content (nonanimal contents are more complex than animal contents; multiple contents are more complex than single contents), and determinants (any determinant other than form is more complex than form alone; multiple determinants are more complex than single determinants). Many R-PAS variables are highly correlated with the Complexity variable, so an option for scoring all the variables adjusts each one based on the underlying Complexity score (for the exact method used, consult the R-PAS manual, p. 303; Meyer et al., 2011). Practitioners should understand the implications of interpreting the raw scores versus interpreting the complexity-adjusted scores. It is recommended that the raw scores more routinely be used, unless Complexity is extremely high or extremely low.

R-PAS: INTERPRETATION

As a reminder, variables in the R-PAS are separated into “Page 1” variables and “Page 2” variables. Page 1 variables have very strong empirical support for their interpretation and, as such, can be used confidently. Page 2 variables have research support, but not as strong as Page 1 variables. The support may not be as strong because the effect size in meta-analyses is not as high, or, more often, because there are simply not many research studies on the variable in question. It is expected that variables will inevitably shift, from Page 1 to Page 2, from Page 2 to Page 1, from not on the list (because of weak empirical support) to on the list, or from on the list to off the list as the R-PAS continues to develop. The following organization and interpretive strategies are based on the current state of the knowledge as of the publication of this book, and they are based largely on the interpretations in the R-PAS manual (Meyer et al., 2011) and the Mihura et al. (2013) meta-analysis as well as Comprehensive System interpretive information that applies to the R-PAS.

Administration Behaviors and Observations

The administration behaviors and observations variables offer some broad context about the engagement of the individual in the process of taking the Rorschach. The behaviors evaluated, though, could be the outcome of a variety of underlying reasons. For example, an individual who turns the cards often may be doing so out of curiosity and interest in the stimulus. However, the reason for doing so may equally as likely be due to anxiety, oppositionalism, paranoia, or a host of other factors. As such, the prompts, pulls, and card turning variables should be interpreted cautiously, and the information from these variables should only be interpreted in the context of both other R-PAS variables and information from outside of the Rorschach testing procedure.

Prompts (Pr)

A high number of prompts signifies that the client needed repeated reminding that he or she should give more than one response per card. Many factors could contribute to

this pattern of responding. Cognitive factors may play a role, with low cognitive ability, cognitive inflexibility, or hypervigilant attention to detail causing difficulty switching mental sets (mentally “resetting” in order to see something new in the same stimulus). Emotional factors, such as depression, anxiety, or fear and mistrust, may also play a role. Finally, personality factors, such as defensiveness, oppositional attitudes, and passive-aggressiveness, may also cause a need for multiple prompts within an R-PAS administration. Each of these may be a hypothesis, but elevation on Pr alone does not support any one of these interpretations specifically.

Pulls (Pu)

Elevations on Pu (a high number of pulls) signify that the client needed repeated reminding that he or she did not need to give more than four responses to any one card. Similar to Pr, elevations on Pu may be caused by multiple different factors. Cognitive expansiveness or confusion may be present, as may emotional issues, such as manic or hypomanic states or anxiety. Positive traits, such as high achievement orientation, may result in elevation on Pu. However, other traits may similarly elevate Pu, such as difficulty following directions, oppositional tendencies, or difficulties with boundaries. Again, elevations on Pu should not be singularly interpreted as any one of these hypotheses, but rather interpreted in the context of the rest of the R-PAS, other tests, and information from outside of testing.

Card Turning (CT)

Like the other observation variables, CT elevations may be caused by multiple different factors, such as mental curiosity and flexibility, oppositional tendencies, obsessiveness and compulsivity, or even being overwhelmed by the task. However, the primary purpose for the inclusion of the CT variable is for help interpreting reflections (r), so direct interpretation of CT should be avoided.

Engagement and Cognitive Processing: Page 1

The Page 1 Engagement and Cognitive Processing variables cover multiple aspects of cognitive complexity/simplicity, ability to cope and adapt to the environment by engaging with the world, and likely reactions to different types of cognitive and emotional stressors. One of the most important variables is Complexity, which serves as an adjustment variable for all the other variables, as it is highly correlated with many of the other R-PAS variables.

Complexity

The Complexity variable is calculated using measures of complexity of the use of location (e.g., with the use of synthesis in a response being more complex than vagueness and using the whole blot or integrating white space being more complex than usual or unusual part areas of the blot), content (e.g., with multiple content codes being more complex than single codes within a response), and determinants (e.g., with multiple determinants being more complex than single determinants within

a response). The variable measures the overall complexity of processing information, as represented by utilizing complex mental abilities in the performance on the R-PAS task. Complexity can have both benefits and risks associated with it. Those high in Complexity interact with the world in a psychologically sophisticated and flexible manner. They are likely to want to engage in more complex problem-solving situations, careers where they are challenged by their environment, and debates and conversations about issues that are not black and white. However, high Complexity may represent cognitive confusion and being overwhelmed, even to the point of manic or psychotic thinking. Trauma-related flooding can result in highly complex profiles as well. At somewhat elevated levels, Complexity may represent anxiety or ruminative thinking and introspection. Practitioners should always be aware that elevations on Complexity may be related to malingering or exaggeration, as those attempting to appear pathological may offer much more complex and complicated responses.

Low scores on Complexity may be related to cognitive or emotional factors. Cognitively, such scores may reflect cognitive simplicity, rigidity, or low cognitive functioning. A simplistic view of the world and simple, straightforward engagement with the environment are reflected in responses that would not elevate the Complexity score (e.g., using single blot area locations, single contents, and a great deal of only form). Emotionally, low Complexity may be caused by emotional withdrawal, secondary to depressive or anxious processes. An anxiety-based reticence to reveal oneself, engage in a task that may expose some vulnerability, or connect relationally may also result in low Complexity. Although trauma histories can lead to flooding and high Complexity profiles, trauma can also lead to emotional numbing and withdrawal, which can lead to low Complexity profiles. Much as high Complexity scores can be situationally related to malingering, low Complexity scores can simply be manifestations of defensiveness or insecurity about how one is performing on the task. It should be noted that when Complexity scores are extremely low, less data and thus less information will be present across the rest of the R-PAS variables.

Number of Responses (R)

The number of responses given by an individual during the R-PAS may be interpreted in many different ways and is generally best interpreted within the context of other R-PAS variables and information from outside of the test content itself. R should always be interpreted with the relative number of Prompts and Pulls performed within an administration. A high R with high Pr would likely occur for very different reasons than a high R with high Pu; for example, a high R, high Pr protocol may relate to some cognitive problems with short-term memory (needing to be reminded with each new stimulus card that more than one response is needed) but high compliance, whereas a high R, high Pu protocol may be related to a manic, overproductive state. Assuming both Pr and Pu are within normal limits, there are some basic possible hypotheses related to both high and low R.

High R may be related to cognitive factors, such as high intelligence and verbal fluency, obsessiveness or perfectionism, or cognitive expansiveness and confusion. Alternatively or in addition to these, high R may be related to emotional factors, such as mania or hypomania or anxiety related to needing to please. Finally, personality variables like needing to be the focus of attention and striving for achievement may result

in high R. Low R may be related to defensiveness, cognitive limitations, or rigidity, similar to low Complexity.

Form % (F%)

The relative degree of using pure form as the defining determinant when compared to the total number of responses (F%), calculated as the number of F determinants divided by R, is a measure of how concrete and simplistic an individual's engagement with the world is. Individuals with high F% have approached the task of the Rorschach in a simplistic manner, using the shape and contours of the blot as the primary driver of responses and ignoring other nuances within the blots, such as the variation in ink tone (shading) or color. Similarly, they are likely to engage in the world in a way that minimizes ambiguity and nuance. This may not always be a negative trait, as it may be effective to cope with ambiguous situations or those with uncertain outcomes by reducing them to basic and simple aspects. However, with more uncertainty and ambiguity in situations, these individuals are likely to struggle to adapt effectively. Those with low F% are able to understand different aspects of situations, looking for nuance and subtle components. They may be more effective in interpersonal relationships, which are necessarily complex in nature. However, they may also have difficulty seeing the big picture if they are too preoccupied by details, contradictions, and subtle components. In general, it is important to interpret F% along with Complexity, R, Blends, and Synthesis (see next sections) to get a clearer picture of how the individual engages with the complexities of the world.

Blends

The use of Blends further signifies the richness and complexity with which an individual perceives and engages with the world around him or her. Individuals high in Blends approach situations with multiple simultaneous perspectives. That is, they tend to take in information and examine it through different lenses, which can help them understand complex and multifaceted situations. However, this rich engagement with the world can be overwhelming and confusing at times. Those low in Blends take a much more unidimensional approach to novel situations and stimuli, attending to a single aspect or component when making judgments. While they may miss some of the complexity of situations, they are likely to be much quicker to arrive at decisions, as they generally have less information to consider.

Synthesis (Sy)

Another variable related to the sophistication and complexity with which an individual engages the world, the amount of synthesis (Sy) used relates specifically to the ability and tendency to integrate ideas and cognitions together. Individuals with high Sy engage in cognitive processes that are sophisticated and interconnecting. Ideas are taken in and incorporated into a more complex world of ideas, interrelating one idea with another. Those low on Sy engage in more linear, straightforward, and obvious thinking patterns, interconnecting ideas much less often. They may have limited cognitive ability, or they may simply have a style of direct, simple, straightforward thinking.

Human Movement and Weighted Color (MC)

Called Experience Actual (EA) in the Comprehensive System, MC is a measure of overall psychological resources and capacity to cope with the world. High MC reflects the ability to fully engage with the environment and world around the individual with effortful, deliberate strategies to cope. These individuals generally adapt to the world effectively, though at times high MC can represent mental activity and engagement put to bad use (e.g., if there are significant perceptual or ideational distortions, high engagement with the world may take on a maladaptive quality). Those low in MC have more limited resources for coping with the world. It should be noted, though, that MC may come out as low due to situational guardedness or defensiveness.

Coping Effectiveness (MC–PPD)

Referred to as D in the Comprehensive System, MC–PPD is a general measure of how effective an individual is at coping with the demands being placed on him or her. It is a means of evaluating the degree of available resources the person has (MC) versus the amount of disorganized events that are occurring beyond the person's control (PPD). Individuals with high MC–PPD are able to handle the current stress demands on them well, using their internal coping resources. They are less vulnerable to disruption during situations that place great pressure on them. Those low on MC–PPD have limits to their ability to handle stress with their internal resources, easily feeling overwhelmed, overloaded, and distracted. They are less able to deal with complex or ambiguous situations, and their thoughts, feelings, and behaviors may be impulsive and poorly focused. Alternatively, highly intellectualized individuals with low MC–PPD may simply currently be engaging in anxious overthinking about existential matters. They will likely perform better in highly predictable situations. It should be noted that MC–PPD tends to be lower on protocols with high Complexity, so it is important to use the Complexity-adjusted scores in these cases.

Human Movement (M)

Human Movement (M) is generally a measure of the bridging of the outside world of reality with internal (mental) action. That is, M assesses how an individual takes the world and elaborates it, thinks about it, and uses it internally in his or her mind. It includes thinking, planning, imagination, and even empathy. M is also an inhibitor of outward behavior, even though that inhibition may only be temporary. Individuals with high M are likely to have good ego functioning, ability to plan, impulse control, and ability to withstand frustration. They may also, however, have an overdeveloped fantasy life. High M may be a reflection of high general (and especially verbal) intelligence, creativity, and abstract reasoning. The quality of this inner life with high-M individuals should be evaluated with other scores related to possible confusion or disturbance in thinking, such as the Cognitive Codes (see the “Cognitive Codes” section earlier in the chapter). Low scores on M may mean several things, including lower intelligence, ideational lethargy related to depression, a lack of imagination, or being the type of person who acts more impulsively rather than thinking through decisions and actions carefully.

M Proportion (M/MC)

The M/MC score is comparable to the EB score in the Comprehensive System, though it functions as a calculated score, and thus dimensionally, rather than a ratio, which is more categorical. However, it similarly assesses the extent to which a person is internally oriented as opposed to being more externally directed and behaviorally responsive to outside stimuli. Related to coping style as well as decision making, the M/MC score assesses whether situations are dealt with in a deliberate, thoughtful manner or tend to be dealt with in a reactive, action-oriented manner. Individuals high on M/MC tend to be more deliberate and thoughtful in their approach to situations, with thoughts and ideas weighing more heavily on decisions than emotions. Much of the literature and the Comprehensive System refer to this style as introversive, though the R-PAS has moved away from this nomenclature. Those low on M/MC are more spontaneous and reactive in situations, relying more on emotions and gut feelings than logic and thinking through. They tend to favor a trial-and-error approach to problem solving, rather than a deliberate, thought-out series of steps. This style used to be called extratensive. Individuals who score in the middle range of M/MC are either guided both by thoughts and feelings in situations, balancing the two often adaptively, or, if MC is low, have low resources for approaching situations in either way.

Color Dominance Proportion [CFC Proportion; $(CF+C)/SumC$]

The Color Dominance Proportion (CFC Proportion) evaluates the general strategies individuals employ when faced with emotionally arousing stimuli and situations. When color dominates, and thus the proportion is higher, reactions are less modulated by thought, careful consideration, or cognitive control. Individuals who score high on CFC Proportion react to emotionally arousing stimuli and situations without much careful cognitive consideration. They may react impulsively with an immediately reactive style. In contrast, those who score low on CFC Proportion are more likely to consider the potential consequences of their actions carefully, thinking through their response to these stimuli and situations. At an extreme, they may be behaviorally reticent throughout their lives, unable to live in the moment and act spontaneously.

Perception and Thinking Problems: Page 1

The Page 1 Perception and Thinking Problems variables cover multiple aspects of potential disturbance in the way one takes in information, processes it, and then thinks about it and uses it as the basis for making judgments. Taken together, they represent mental health and psychopathology related to thinking and thought-related problems.

Ego Impairment Index-3 (EII-3)

The Ego Impairment Index-3 (EII-3) combines variables related to disturbance in thought *processes* (FQ–, WSumCog), including problematic social cognition and interpersonal comprehension (M–, GHR, and PHR), and disturbing thought *content* (Critical Contents). As such, it is a general measure of thought-based psychopathology (such as, at an extreme level, psychosis). Individuals who are high on EII-3 have both disturbed thinking processes and contents, with higher elevations generally

indicative of a higher likelihood for serious psychopathology. Those low on EII-3 have linear thought processes, good reality testing, and a low degree of intrusive disturbing ideation.

Thought and Perception Composite (TP-Comp)

The Thought and Perception Composite (TP-Comp) is similar to the EII-3, except that it focuses entirely on thought *process* and not content. It is calculated with a combination of variables related to reality testing (FQ-%, WD-%, and M-) and those related to confusion and disorganization of thoughts (WSumCog and FAB2). Individuals with high TP-Comp scores (raw score > 2) have problems perceiving the world accurately and realistically, as well as thinking clearly and logically. Extreme elevations (raw score > 3.5) suggest severe thought disturbance, most often associated with psychosis. Individuals who score low on TP-Comp (raw score < 0.5) perceive the world accurately and think in a logical and clear way. This kind of logical, accurate thinking and perception may be associated with simplicity and straightforwardness (it can be useful to interpret this variable along with Complexity), or it can be related to a healthy but sophisticated way of engaging with mental life (when accompanied by high Complexity).

Weighted Sum of the Six Cognitive Codes (WSumCog)

The Weighted Sum of the Six Cognitive Codes (WSumCog), known in the Comprehensive System as WSum6, is a measure of cognitive distortion and thought disturbance. Individuals who score high on WSumCog have a serious difficulty with reality, strained reasoning, faulty cause-and-effect relationships, loose associations, disorganized thinking, and poor ability to focus on tasks. Those with slight elevations may have immature thinking or a playful engagement with the Rorschach test itself, which can be examined and gleaned from the responses themselves. For example, if a great deal of the responses are fanciful but somewhat reasonable, responses that include Level 1 INC or FAB codes (such as horses with wings or bears playing patty cake), interpretation may focus more on immaturity or playfulness. Alternatively, if the elevation is based on responses that include multiple DR, especially at Level 2, or PEC codes, then interpretation may be more aligned with thought disturbance. Those with low WSumCog tend to think in a more conventional manner, which could be healthy and adaptive or, at very low levels, rigid and concrete.

Severe Cognitive Codes (SevCog)

One way of evaluating the seriousness of WSumCog is by comparing it to the Severe Cognitive Codes (SevCog) variable, which is similar but only includes Level 2 of DV, DR, INC, and FAB, in addition to any PEC and CON codes. Individuals who score high on SevCog are much more likely to have thought disturbance, with high elevations related to psychotic thought processes. As with WSumCog, though, evaluating the actual content and reason for the codes may reveal that playfulness or dramatic flair was the cause for elevation, rather than confusion or idiosyncratic thinking. Individuals who score low on SevCog, similar to WSumCog, tend to think in a more conventional manner, which could be healthy and adaptive or, at very low levels, rigid and concrete.

FQ– Percentage (FQ–%)

The FQ– Percentage (FQ–%), called the X–% in the Comprehensive System, is a measure of the degree to which a person has distorted perceptions of reality. The higher the FQ–%, the more likely that the person has a significant level of impairment relating to distorted perceptions. High percentages are found in depression, and extreme elevations are often found in schizophrenia. Those who score high on FQ–% have distortions in the way they view reality, such that their behaviors, judgment, and reactions to situations are often problematic. Individuals who score low on FQ–% have more conventional and accurate perceptions of reality.

WD– Percentage (WD–%)

The WD– Percentage (WD–%) is similar to the FQ–%, in that it is a measure of distortion in the perception of reality; however, it focuses only on distortion when conventional blot locations are used, which constitute situations where conventional perception of reality is likely more easily accomplished. That is, the accumulation of FQ– codes on locations that are W or D (measured by WD–%) are even more problematic and atypical than FQ– codes on Dd areas (which are included in the FQ–% code). Interpretively, WD–% is fairly similar to FQ–%, such that individuals who score high on WD–% have disturbances in reality testing, most often related to thought-based psychopathology. Those who score low on WD–% have more conventional and realistic reality testing capabilities. The R-PAS manual (Meyer et al., 2011) suggests that looking at WD–% and FQ–% together is the best measure “to assess perceptual judgment and reality testing” (p. 360).

FQo Percentage (FQo%)

The FQo%, conceptually and computationally similar to the X+% on the Comprehensive System, is an indicator of the degree to which a person perceives things in a conventional, realistic manner. Individuals who score high on FQo% generally have good reality-testing abilities and are, for the most part, adaptive and psychologically healthy. Those who score low on FQo% may do so for two primary reasons. First, if FQo% is low and FQ–% and WD–% are high, then the individual has perceptual difficulties and problems with understanding reality in a logical way. If FQo% is low and neither FQ–% nor WD–% is elevated, it more likely represents a unique, idiosyncratic way of interacting with the world.

Popular (P)

The number of Popular (P) responses reflects the respondents’ degree of similarity to most people (especially the way they perceive the world), the extent to which they conform to social standards, and the relative ease with which they are likely to be influenced in interpersonal relationships. Individuals who score high on P have conventional ways of viewing the world, but those with excessive elevation on P tend to have stereotyped and rigid views of the world. Low scores on P are harder to interpret. Multiple factors can cause low P. Reality testing problems may cause low P (check FQ–% and WD–%), as can uniqueness and individuality in approaching the world

(see the last section, “FQo Percentage [FQo%]”) or simply deliberately suppressing Popular responses, perhaps as a specific approach to taking the Rorschach based on conceptions of how common responses may be interpreted by the examiner.

Stress and Distress: Page 1

The Page 1 Stress and Distress variables focus primarily on themes of current distress; however, they also include some information on some traits related to how an individual deals with stressors in everyday life. These variables should be considered within the context of all other variables, as situational distress may affect current performance on the R-PAS. Additionally, variables in other domains may illuminate some causes or reinforcers for current distress.

Sum of Shading and Achromatic Color (YTV C')

The Sum of Shading and Achromatic Color (YTV C') variable, measured by simply adding the number of each of these codes to have been applied, is one of the more trait-like variables in this cluster and provides some context for why and how an individual may be experiencing or not experiencing situational distress. The YTV C' variable relates to a tendency to focus on nuance, uncertainties, and ambiguity in the world. Individuals who score high on YTV C' may not necessarily experience distress. These individuals may simply be intellectually and emotionally drawn to nuance, detail, and even small fluctuations in their emotional world, their interpersonal relationships, and their broader context. However, those who score high on YTV C' may also find all the uncertainty on which they focus their attention quite unnerving and overwhelming. Those who score low on YTV C' tend to focus on the big picture much more than on nuance or detail and deal with situational stressors in a straightforward manner.

Inanimate Movement (m)

Inanimate Movement (m) provides an index of the extent to which persons are experiencing life events, often interpersonal, that are beyond their ability to control, causing mental agitation and cognitive anxiety. Individuals who score high on m have anxious thoughts in the face of whatever situational stressors are present. Depending on what is actually going on in a client's life, this anxiety may be entirely expectable and realistic. Those who score low on m are either not facing significant stressors currently, or if they are, these stressors are not causing significant conflict and tension within the individuals.

Diffuse Shading (Y)

Diffuse Shading (Y) represents a sense of helplessness and withdrawal, which is frequently accompanied by anxiety and is often a response to ambiguity. Individuals who score high on Y are currently experiencing psychological pain and have resigned themselves to their situation. Similar to m, the helplessness associated with high Y may be expected and realistic in the face of whatever real stressors an individual is facing. Those who score low on Y are either not currently experiencing any significant stressors or ambiguity that would cause a sense of helplessness, or if they are, the individuals are not responding to them with a sense of helplessness.

Morbid Content (MOR)

Morbid Content (MOR) relates to general pessimism, negative self-image, and generally negative ideation. It is important to note that while these components are often present in depression, they are also present in many other forms of psychopathology. As such, the MOR scale more reliably differentiates those who have psychopathology from those who do not than distinguishing depression itself from other forms of psychopathology (Meyer et al., 2011). Individuals who score high on MOR feel damaged or harmed by the world and life in general in some way. They may have a general underlying sense of dysphoria. Those who score low on MOR are more likely to be optimistic and self-assured, with few insecurities around being abnormal or flawed in fundamental ways.

Suicide Concern Composite (SC-Comp)

The Suicide Concern Composite (SC-Comp) is a measure for the potential for suicidality and genuine self-harm, and it is not related to suicidal or self-harm gestures without the actual intent to harm oneself seriously. It is calculated using multiple variables, including current distress variables (MOR), sensitivity to stress (CBlend), coping abilities (MC–PPD, CFC Proportion), level of complexity in thinking and engagement with the world (Complexity, SI, V, FD), and other facets, such as self-focusing, oppositionality, realistic appraisal of the world, and others (Symmetry, SR, FQo%, P, H). The calculated resulting SC-Comp variable has been associated with heightened risk for suicidality. Individuals who score high (raw score around 7, and especially above 8) are at risk for suicide, though this risk may not be evident, even to the clients themselves. Those who genuinely attempt or commit suicide elevate this scale, but not all those who elevate this scale go on to attempt suicide. As such, it is important for clinicians to be sensitive to this fact, not to overreact or catastrophize a situation in which SC-Comp is elevated, but also to engage the client in meaningful dialogue about potential risk for self-harm.

Self and Other Representation: Page 1

The Page 1 Self and Other Representation variables cover multiple aspects of how a client perceives and thinks about him- or herself, others, and relationships in general. Interpersonal styles (such as dependency and oppositionality) are included, as are possible distortions in how a person views him- or herself and others and expectations for interactions with others. As with all the other clusters of variables, these should be considered within the broader context of all the R-PAS variables, as well as with information from outside of the R-PAS, as many factors (such as current distress, thought disturbance, etc.) can interact with how individuals view themselves and others.

Oral Dependency Language (ODL%)

Oral Dependency Language (ODL%) has replaced the simple use of food (which is the primary measure of dependency in the Comprehensive System) as a more empirically supported measure of implicit dependency needs. Individuals who score high on ODL% are likely to display behaviors (even though they may not always be overt)

related to needing nurturance and support from others, including eliciting leadership and assertiveness from others through their own passivity. They are sensitive to rejection and need assurance of others' commitment to them. Those who score low on ODL% are extremely independent and have no problem being assertive in interpersonal situations.

Space Reversal (SR)

Space Reversal (SR) is a measure associated with negativism and oppositional tendencies as personality variables, as opposed to anger-related oppositionalism in any one situation. That is, SR measures the tendency to go against the grain and oppose what is normatively expected, which may be related more to a fiercely independent and individualistic style of interacting with the world or a disdain for rules. Individuals who score high on SR are oppositional and prefer to follow their own rules, rather than those imposed upon them. Those who score low on SR are likely to be more compliant, obedient, and respectful of rules, norms, and "appropriate" behavior in social and other situations.

MAP Proportion (MAP/MAHP)

The MAP Proportion (MAP/MAHP) variable, which should only be interpreted confidently when at least four Mutuality of Autonomy (Pathology or Health) codes have been applied within a protocol, represents how healthy an individual's object representations are. An individual with healthy object representations can view him- or herself in realistic, healthy ways and expect positive, mutually respectful and healthy interactions with others. An individual with unhealthy object representations may have difficulty viewing and understanding him- or herself realistically and expect others to be motivated by negative intentions; to behave in ways that will exploit, let down, or harm the individual; and to generally be not trusted or relied on. As such, individuals who score high on MAP/MAHP have distorted and unhealthy, unrealistic views of themselves and others. Those who score low on MAP/MAHP have more realistic, mature views of themselves and others, expecting relationships to be positive, mutually beneficial, and generally more healthy and reciprocal.

PHR Proportion (PHR/GPHR)

The PHR Proportion (PHR/GPHR) measures general social skills and healthy normative social interactions, in contrast to the MAP Proportion, which more closely measures the capacity for interpersonal closeness. Similar to the MAP Proportion, the PHR Proportion should be interpreted confidently only if at least four Human Representations (Good or Poor) have been coded within a protocol. Individuals who score high on PHR Proportion have a poor understanding of themselves and other people, which in turn will lead to problematic interpersonal interactions. The quality of the distorted views of people can be evaluated by looking at the specific content of the responses coded as PHR, which may include representations of humans as damaged, aggressive, distorted, or many other poor qualities. Those who score low on PHR Proportion are more skilled and competent in interpersonal interactions, based on an

adequate understanding of themselves, others, and normal social interactions. It can be important to interpret this proportion together with the MAP Proportion, as PHR Proportion looks more at social skills and understanding of appropriate social behavior, while MAP Proportion looks more at the capacity for closeness with others.

Human Movement Minus (M–)

The Human Movement Minus (M–) variable, which looks at responses that include M as a determinant and have form quality minus, functions as a measure of the capacity to mentalize (see Bateman & Fonagy, 2012) with others, which includes realistically appraising others' intentions, motivations, and behaviors. Individuals who score high on M– have a distorted and inaccurate understanding of how others think and why they behave the way they do. As such, their interpersonal relations are disrupted, as they tend to be ineffective in interpersonal interactions and relationships. Those who score low on M– have better capacity to mentalize with others, more realistically understanding their intentions, attitudes, thoughts, and behaviors. The R-PAS manual (Meyer et al., 2011) reports that longer and more complex protocols can produce false positives on this variable.

Aggressive Content (AGC)

Aggressive Content (AGC) is a relatively straightforward measure of aggressive tendencies or competitiveness. However, as a highly face valid variable (i.e., respondents can easily guess that including aggressive content will likely be interpreted as anger or aggression), individuals can easily omit aggressive content, making low scores on AGC less interpretable than most scales. Individuals who score high on AGC, though, have a style that is either outright aggressive (which assumes some behavior that does not respect others and their rights) and oriented toward power or competitive and focused on high achievement (which is a healthier and more socially acceptable form of aggression). Other variables, and especially information from outside of the test and testing situation, can help practitioners distinguish between these two interpretations of elevations on AGC.

Whole Human Content (H)

Whole Human Content (H) represents individuals viewing themselves and others as whole, integrated people, which is generally a healthy capacity for understanding themselves as well as interacting with others in healthy ways. Individuals who score high on H understand themselves and others as complex, multidimensional beings, irreducible to simple identifiers. Those who score low on H tend to view others in more basic, simple terms, such as reducing them to their social status or role (e.g., seeing all women or men as somehow the same) or reducing them to oversimplified objects to be feared or disliked. These individuals tend to have poorer interpersonal relations as a result of their difficulty seeing people as whole, integrated beings.

Cooperative Movement (COP)

Cooperative Movement (COP) is one of the R-PAS variables that looks specifically at a positive and prosocial trait within the client. COP is a measure of interpersonal

expectancy, such that individuals who score high on COP tend to expect that social interactions and relationships will be positive, supportive, and collaborative. Like MOR, COP has high face validity, such that respondents can easily guess that including responses with cooperative movements and relationships in them is likely to cast them in a favorable light. As such, it is important to interpret COP with this in mind and to look to other variables, tests, and information to confirm the hypotheses that emerge from elevations on COP. Those who score low on COP either have little interest in social interaction and relationships, or if they do, they do not tend to expect that they will be supportive and rewarding. Low scores may not be dispositional, however, as those with significant current distress may not produce many COP responses.

Mutuality of Autonomy Health (MAH)

Like COP, the Mutuality of Autonomy Health (MAH) variable looks explicitly for a positive and prosocial trait, the ability to create and sustain mature, healthy, genuine, and intimate relationships with others. Also like COP, however, the type of response that are coded MAH are quite face valid, such that respondents wanting to be seen as healthy can easily guess to include such responses. Regardless, individuals who score high on MAH tend to have the capacity to form close, intimate, and genuine relationships with others. Those who score low on MAH may not necessarily have difficulty with intimacy and closeness, but they may; the MAP Proportion score can help determine if this is the case.

Engagement and Cognitive Processing: Page 2

There are many Page 2 Engagement and Cognitive Processing variables, which cover multiple aspects of cognitive sophistication, flexibility, and engagement with arousing and stimulating situations. As Page 2 variables, interpretation of each should be more tentative than with Page 1 variables, and practitioners should rely on more extreme scores for interpretation. The R-PAS manual (Meyer et al., 2011) recommends interpreting these variables only if they are above a standard score of 115 or below a standard score of 85.

Whole Percentage (W%)

The Whole Percentage (W%) variable, which looks at the percentage of all the responses in a protocol that used the entire blot, reflects processing and problem solving that is holistic, global, and often abstract. However, global and holistic processing and problem solving can have different qualities. For example, it can be simplistic and reductionistic, which would likely be the case with a low Complexity score and a high Vg% score (see the “Vagueness Percentage [Vg%]” section below). Conversely, it can be sophisticated and involve synthesizing multiple disparate pieces of information into a coherent whole, which would be the case with high Sy and Blends. However, even with complexity and sophistication in processing information into coherent wholes, the process may be marred by poor logic (look at the WSumCog variable) or perceptual inaccuracies (look at the FQ-%, WD-%, and FQo% variables). Thus, individuals who score high on W% tend to process information in holistic ways,

though the exact quality of this processing may be sophisticated or not and accurate and logical or not. Those who score low on W% may get preoccupied by details and have difficulty viewing situations more holistically.

Unusual Detail Percentage (Dd%)

Unusual Detail Percentage (Dd%), which is the percentage of the total responses that use a Dd location code, reflects detail orientation and the tendency to look at situations in an idiosyncratic and even bizarre manner. Individuals who score high on Dd% attend to small details of situations and use these details to make larger judgments, which is not a typical way of interacting with the world. This may simply reflect detail orientation, but it may also be related to obsessiveness and even paranoia, in which individuals may perceive small signs and attribute to them large, negative meanings. Those who score low on Dd% tend not to focus on unusual aspects or components of problems, situations, or behaviors, and they likely use more conventional information to form judgments.

Space Integration (SI)

Space Integration (SI) represents a very different conceptual meaning than Space Reversal (SR), and the R-PAS has separated these two variables out because of this fact. Integrating space into a larger response requires complex cognitive processing, related to attending to and integrating multiple sources of information into a coherent whole. Individuals who score high on SI are flexible in their thinking and take in and use multiple sources of information in order to form judgments and make decisions. Those who score low on SI may not necessarily be simplistic in their thinking, but low scores on SI should be considered alongside low Sy and low Complexity scores to determine the level of sophistication and complexity of thinking the individual typically uses.

Intellectualized Content (IntCont)

Intellectualized Content (IntCont) is calculated in the same way as the Intellectualization Index on the Comprehensive System, as a sum of 2 times the coded abstract (Ab) content plus any Art and anthropology (Ay) codes. The IntCont variable itself represents the tendency to intellectualize, which includes using analytic perspectives to distance oneself from emotions, including social discomfort. Individuals who score high on IntCont tend to neutralize emotions or social discomfort through analyzing things from an intellectual perspective, denying or concealing the impact of emotions. They are generally pretty successful at mitigating the impact of emotions through pseudo-intellectual means, but situations that cause extreme levels of emotion are likely to overwhelm and disorient them, as they are not used to or good at dealing with emotions. Those who score low on IntCont do not employ an intellectualizing defense mechanism to deal with uncomfortable emotions or situations, though they may not necessarily be more comfortable than those who score high on IntCont with their emotions, as they may employ other mechanisms for distancing themselves from them.

Vagueness Percentage (Vg%)

The Vagueness Percentage (Vg%) is calculated as a percentage of the total number of responses that are coded with vague (Vg) object quality and represents an immature and unsophisticated ability to analyze and synthesize information. Individuals who score high on Vg% may be cognitively limited, impulsive, defensive, or simply unskilled in synthesizing information in a sophisticated and nuanced manner. However, when Vg% and Sy% are both elevated, it suggests inconsistency in the way the individual processes information, which may be situationally influenced. Those who score low on Vg% have a more sophisticated way of analyzing information, though just how sophisticated their skills are should be evaluated with the Sy% and Complexity variables alongside Vg%.

Vista (V)

The Vista (V) code is listed twice in the R-PAS, as it takes on two slightly different meanings that are related both to the Engagement and Cognitive Processing domain and the Stress and Distress domain. With respect to the Engagement and Cognitive Processing domain, V reflects the ability and tendency to take perspective and look at situations from different vantage points, including scrutinizing potentially hidden or masked aspects of situations. As such, individuals who score high on V are able to distance themselves from situations and evaluate them from a different perspective. This may be a healthy and sophisticated way of approaching situations, though at extremes it may represent dissociation or viewing oneself or others as globally negative. Those who score low on V may not necessarily have difficulty taking perspective, though this may not be their primary mode of approaching information in the world.

Form Dimension (FD)

Like V, the Form Dimension (FD) code also represents a form of perspective taking. However, while V incorporates taking perspective on situations by looking through things or scrutinizing potentially hidden aspects, FD focuses specifically on creating distance from a situation in order to judge it. Individuals who are high on FD distance themselves from situations in order to take a different, often wide-angle perspective on them. At extreme elevations, this distancing may represent reactions to trauma. Similar to those who score low on V, those who score low on FD may not necessarily have problems with perspective taking, but they may prefer to stay close to or deeply within situations when trying to figure them out.

Response to Final Three Cards Percentage (R8910%)

Although there is limited research to support this variable, the Response to Final Three Cards Percentage (R8910%) seems to measure how responsive individuals are to compelling, potentially emotionally arousing situations, as the final three cards are vibrant and have multiple colors. The variable is calculated as a percentage of the total number of responses on a protocol that are given on the last three cards (VIII, IX, and X). Individuals who score high on R8910% are intrigued by and engage fully with situations

that are compelling, including emotionally arousing interpersonal situations. Those who score low on R8910% may avoid or shrink away from these same situations, perhaps finding them overwhelming or unpleasant.

Weighted Sum of Color (WSumC)

Color does not seem, as previously believed, to be directly related to the emotional experiences of an individual. Rather, the Weighted Sum of Color (WSumC), which calculates the sum of color determinants, weighted by their proportional use of form (C is weighted more heavily than CF, which is weighted more heavily than FC), reflects how interested individuals are in the saliently compelling features of everyday life, which may include emotionally arousing features. WSumC is also considered a positive, adaptive component of coping with life. Individuals who score high on WSumC are lively, vital, and engaged in exciting aspects of the world around them, including their emotional world, and they may behave in reactive ways. Those who score low on WSumC are less attuned to or interested in the exciting, vibrant aspects of life, often preferring routine and predictability.

Pure Color (C)

Pure Color (C), unmodulated by form, reflects a general tendency to judge and react to situations immediately and fully, without the use of thinking to temper reactions and gut instincts. Individuals who score high on C trust their gut instincts and immerse themselves in their experiences fully. This style may be problematic if it is related to impulsivity, emotional overreactivity, or haste (other information from the R-PAS and outside of the test can help determine if these qualities are prevalent). However, this style may also be adaptive, as it relates to fully immersing oneself in positive experiences and truly enjoying them, without second guessing or questioning their value. Those who score low on C are more likely to modulate and evaluate their gut instincts with thinking and logic (whether good or poor logic) before making judgments and acting.

Mp Proportion [$Mp/(Mp+Ma)$]

As a proportion of passive human movement to all human movement (active or passive), the Mp Proportion reflects a general tendency toward a passive stance in dealing with situations (including retreating into fantasy or ruminating) or interpersonal passivity. It should be interpreted only if the total number of human movement (M) codes is at least 4. Individuals who score high on Mp Proportion tend to withdraw into thought (and rumination) or fantasy in situations that require action, and they may take a passive stance interpersonally. Those who score low on Mp Proportion are less likely to retreat into their thoughts and may take a more active role in solving problems and making decisions.

Perception and Thinking Problems: Page 2

There is only one Page 2 Perception and Thinking Problems variable. Again, as a Page 2 variable, interpretation should be more tentative than with Page 1 variables, and practitioners should rely on more extreme scores for interpretation. The R-PAS manual

(Meyer et al., 2011) recommends interpreting this variable only if it is above a standard score of 115 or below a standard score of 85.

FQu Percentage (FQu%)

The FQu Percentage (FQu%), which is the least reliable of the form quality scores and is thus on Page 2 interpretively, represents unconventional and unique ways of interacting with and interpreting the world. Individuals with high FQu% view the world in a unique and interesting way, though it is likely not as potentially dysfunctional as FQ-%, which often represents actual misinterpretations of the world. While individuals high on FQu% may not approach the world in a conventional and usual manner, the way they interpret situations is most often at least logical. Individuals who score low on FQu% may be more conventional in the way they perceive the world (especially if FQ-% and WD-% are low and FQo% is high), or they may simply have even more problematic perceptions and interpretations of the world than FQu measures (if FQ-% and WD-% are high).

Stress and Distress: Page 2

There are many Page 2 Stress and Distress variables, which, like the Page 1 Stress and Distress variables, relate both to traits and characteristics related to dealing with stressors and the world and to current distress. As Page 2 variables, interpretation of each should be more tentative than with Page 1 variables, and practitioners should rely on more extreme scores for interpretation. The R-PAS manual (Meyer et al., 2011) recommends interpreting these variables only if they are above a standard score of 115 or below a standard score of 85.

Potentially Problematic Determinants (PPD)

The Potentially Problematic Determinants (PPD) variable is calculated using nonhuman movement (FM and m) and shading and achromatic color (Y, T, V, and C') determinants and reflects demands on psychological resources in terms of stimulating and irritating needs, wishes, thoughts, and feelings. As a composite score that has many elements, the PPD does not specify the exact nature of upsetting psychological demands. Elevations in long protocols with high Complexity may suggest depth and sensitivity in the experience of the world. Individuals who score high on PPD typically have uncontrolled demands on their psychological resources, such that they are thinking and feeling things that are upsetting. Those who score low on PPD do not have significant current demands on their mental resources, either because of generally low-stressor situations or because they do not tend to attend to the multiple needs, wishes, thoughts, and feelings that may upset others.

Color Blended with Shading and Achromatic Color (CBlend)

The Color Blended with Shading and Achromatic Color (CBlend) variable looks at responses that blend color (FC, CF, or C) with some sort of shading or achromatic color (Y, T, V, or C') determinant and reflects an aspect of emotional sensitivity and

reactivity. Typically these blends relate to emotional reactions to stimuli being colored by negative concerns, including negative attribution to ambiguity or hesitation because of indefiniteness. That is, even good situations that elicit an emotional reaction are met with hesitation, as negative thoughts can taint the positive stimulation. Individuals who score high on CBlend tend to have their immediate emotional reactions to situations, even good situations, colored by distressing assumptions or discomfort. Those who score low on CBlend do not often experience these mixed emotional states in which negative inferences can spoil even positive situations.

Achromatic Color (C')

Achromatic Color (C') relates both to dark and negative imagery (including the “dark side” of situations) and to attempts to suppress emotional reactivity to situations. Although C' is not related directly to depression, it can reflect an avoidance of emotion, both positive and negative. Individuals who score high on C' are likely working hard to avoid or suppress their emotional reactions to life, often uncomfortable with emotions because they tend to focus on the negative aspects of situations. Those who score low on C' are less likely to suppress their reaction to emotionally arousing situations and may focus more on positive aspects of them.

Vista (V), Revisited

As stated in the Engagement and Cognitive Processing section above, V reflects the ability and tendency to take perspective and look at situations from different vantage points, including scrutinizing potentially hidden or masked aspects of situations. This perspective taking and taking into account potentially obstructed, hidden, or masked aspects of situations can be an uncomfortable process when directed at evaluating oneself or other people. Specifically, a nuanced and complex view of oneself or others will likely uncover both positive and negative attributes, which can lead to some discomfort. For some individuals who are depressed or have other psychopathology, the self-scrutiny associated with high V takes on a self-critical flavor. Individuals who are high on V evaluate themselves and others in a complex, multiple-perspective-taking way, often uncovering hidden or nuanced aspects of themselves and others. They may become highly self- or other critical, which can be distressing. Those who score low on V do not tend to take as sophisticated view of themselves or others, which can be related to reactions less modulated by thought (look at C as well), limited cognitive ability, or simply a more straightforward approach to evaluating situations, self, and other people.

Critical Contents (CritCont%)

The Critical Contents (CritCont%) variable is calculated as a percentage of the total number of responses that include imagery representing thoughts that are often kept personal and hidden from normal social interaction, including MOR, AGM, An, Bl, Ex, Fi, and Sx. The CritCont% has three different potential meanings, related to three different lines of research that support nearly identically calculated scores: CritCont%, the Trauma Content Index (TCI), and Dramatic Contents, which is a measure of potential malingering. What makes interpreting CritCont% so difficult is that it could

validly represent any of the three underlying meanings. The CritCont% research itself focuses on primitive and problematic thought content that can be related to a failure to censor disturbing thoughts. The TCI research focuses on a history of trauma and the tendency to dissociate from emotions and situations. The Dramatic Contents research focuses on exaggeration of disturbing imagery in order to dramatically convey problems, as in malingering, and is related to the high face validity of this particular variable; that is, a respondent may guess that including a great deal of morbid, aggressive, explosive, or sexual content will be interpreted as psychopathological. Thus, individuals with high CritCont% scores may be exaggerating problems and malingering; may have a history of trauma and a tendency to dissociate; or may be dealing with primitive and disturbing thought contents, with extreme elevations potentially related to psychotic thoughts. Those who score low on CritCont% do not likely warrant any of the three preceding interpretations; they are likely not malingering, they likely do not have significant trauma that has caused a tendency to dissociate, and they are not struggling with primitive and disturbing thoughts.

Self and Other Representation: Page 2

The many Page 2 Self and Other Representation variables, like the Page 1 Self and Other Representation variables, relate to varied aspects of one's self-image, perception of others, and interest in interpersonal relationships. As Page 2 variables, interpretation of each should be more tentative than with Page 1 variables, and practitioners should rely on more extreme scores for interpretation. The R-PAS manual (Meyer et al., 2011) recommends interpreting these variables only if they are above a standard score of 115 or below a standard score of 85.

All Human Content (SumH)

The All Human Content (SumH) variable, known in the Comprehensive System as the Interpersonal Interest variable, is a measure of the degree to which a person is interested in other people. Individuals who score high on SumH are highly interested in and acutely aware of other people, which can be healthy and adaptive or can take a hypervigilant or paranoid quality. Other variables on the R-PAS (such as the V-Comp) and information from outside of the R-PAS can help determine if the high attunement to others has any of these negative intentions. Those who score low on SumH have very little interest in interpersonal relationships with others, which may be a more long-standing trait or may be more situational, as other concerns may be preoccupying the client's mind at the moment, such that interpersonal relationships are taking a less important role.

Non-Pure H Proportion (NPH/SumH)

The Non-Pure H Proportion (NPH/SumH) is calculated as a proportion of the total amount of all types of human content [H, Hd, (H), and (Hd)] that are not pure H, and it should be interpreted only if there are a minimum of 4 human content codes (of any type) in a protocol. The NPH/SumH variable reflects how likely an individual is to either perceive or think about him- or herself and others in unrealistic or

fantastical ways. Individuals who score high on NPH/SumH tend not to view themselves and others in a mature, complex, sophisticated way that reflects an accurate representation of who they are. In contrast, they imagine themselves and others in more unrealistic terms, often expecting unrealistic reactions and behaviors from others. It should be noted that at times elevation on NPH/SumH simply reflects a playful or fanciful quality to the inner life of the individual, rather than a process of distortion or misunderstanding. Those low on NPH/SumH are more realistic and mature in their understanding of themselves and others.

Vigilance Composite (V-Comp)

The Vigilance Composite (V-Comp) is calculated using Complexity, oppositionalism (SR), interpersonal interest (T), views of other people (SumH, parenthesized content, whole to partial content), a content code typically related to hiding parts of oneself (Cg), and effortful cognitive functioning (SI). It reflects vigilance, including effortful and deliberate focus on and assumption of potential danger, especially interpersonally, as well as interpersonal distancing and guardedness. Individuals who are high on V-Comp (raw score > 4.5, and especially if raw score > 5.5) are rigidly guarded and distancing from others, constantly wary of others' intentions and motives. They often attend to minor details and attribute malicious or aggressive meanings to them. These individuals should also be evaluated for how realistic and accurate their perceptions of reality are (see FQ-% and M-) and how logical their thinking is (interpret along with WSumCog). There is no research on interpreting low scores on V-Comp, though these individuals are likely not as vigilant and wary of others as those who score high on V-Comp.

Reflections (r)

Reflections (r) on the R-PAS reflect a degree of self-absorption and self-centeredness, needing support and reassurance from others, though this interpretation should be made with special caution as responses that have landscapes reflected do not support this interpretation and thus produce false positive results on r. Humans and animals reflected in a response are more consistent with this interpretation of narcissistic need. Individuals who score high on r need a great deal of reassurance and "mirroring support" (Meyer et al., 2011, p. 374) and, as such, often present themselves in a way that tries to garner attention and affirmation. Extreme elevations may relate to narcissistic tendencies, claiming greatness in the world in order to get positive feedback and affirmation. This need for affirmation may cause affective or interpersonal difficulties if persons do not receive external validation. Those who score low on r do not use themselves as the point of reference when engaging with the world and likely do not have as much need for validation from others.

Passive Proportion [$p/(p+a)$]

The Passive Proportion [$p/(p+a)$] variable is calculated as the proportion of total responses with movement determinants (active or passive) that are specifically coded as passive and should be interpreted only if at least 4 movement determinants are coded

within a protocol. The $p/(p+a)$ variable is a broad measure of passivity in attitudes, values, and behaviors. Individuals who score high on $p/(p+a)$ are more likely to be passive and need others to take the lead and to rely on luck or fate to lead the decisions in their lives. This style is less effective at helping individuals cope with stressors than an active style is. Those with a low $p/(p+a)$ score are more active in their problem-solving strategies and are thus more adaptive in general at coping with stressors.

Aggressive Movement (AGM)

Aggressive Movement (AGM) is a highly face-valid code (i.e., individuals trying either to appear pathological or to suppress aggressive tendencies can easily guess that including more or suppressing aggressive movement, respectively, will achieve these goals) related to aggressive intentions, either from within the respondent him- or herself or attention paid to the person in the environment. It is not as robust as AGC but can be interpreted alongside it, somewhat more tentatively. Individuals who score high on AGM are likely either aware of aggressive intentions in others or have aggressive intentions themselves. Those who score low on AGM may not have the awareness of aggressive intentions, the aggressive intentions themselves, or may be suppressing these in order to look more favorable on the R-PAS.

Texture (T)

Texture (T) relates to the need for supportive interpersonal relationships and closeness. Individuals who score high on T (raw score > 1) have an intense need for affection and closeness and may get significantly upset if their interpersonal needs are not met, which can happen easily if their needs for support are unrealistic. Interpretation for those who score low on T is not as clear, though they likely do not have as intense or strong needs for interpersonal closeness as those who score high on T, or their needs are being met.

Personal Knowledge Justification (PER)

Personal Knowledge Justification (PER) reflects some level of insecurity in the way an individual comes across and a resultant defensiveness or authoritarian stance in which interaction with the world is colored by justification of values, attitudes, and behaviors based on his or her own personal experience. However, it is important to look at the specific content coded PER to see if it is more a sharing of information with the examiner in order to connect, rather than attempts to justify responses. Individuals who score high on PER assert their own personal knowledge and authority to justify their values, attitudes, and behaviors. Those who score low on PER are likely either more secure or more flexible in their attitudes and values, such that they do not feel the need to justify them.

Anatomy (An)

The use of Anatomy (An), which encompasses both An and X-ray (Xy) from the Comprehensive System, reflects preoccupation on the functioning and vulnerability of one's body. Individuals with high An have bodily concerns, functioning, and vulnerability on

their mind. Those with low An scores do not have these same bodily concerns occupying their mental energy.

RECOMMENDED READING

- Exner, J. E. (2000). *A primer for Rorschach interpretation*. Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2001). *A Rorschach workbook for the comprehensive system* (5th ed.). Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2003). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (4th ed.). Hoboken, NJ: Wiley.
- Exner, J. E., & Weiner, I. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents* (2nd ed.). New York, NY: Wiley.
- Masling, M. (2006). When Homer nods: An examination of some of the systematic errors in Rorschach scholarship. *Journal of Personality Assessment*, 87, 62–73.
- Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual*. Toledo, OH: Rorschach Performance Assessment System.
- Mihura, J. L., Meyer, G. J., Bombel, G., & Dumitrascu, N. (2015). Standards, accuracy, and questions of bias in Rorschach meta-analyses: Reply to Wood, Garb, Nezworski, Lilienfeld, and Duke (2015). *Psychological Bulletin*, 141(1), 250–260.
- Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin*, 139(3), 548–605.
- Wood, J. M., Garb, H. N., Nezworski, M. T., Lilienfeld, S. O., & Duke, M. C. (2015). A second look at the validity of widely used Rorschach indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013). *Psychological Bulletin*, 141(1), 236–249.

SCREENING FOR NEUROPSYCHOLOGICAL IMPAIRMENT

An important role in clinical practice is screening and assessing for the presence of possible neuropsychological impairment. The importance of this function is highlighted by data indicating that 20% to 30% of assessment referrals to professional psychologists relate to information regarding the central nervous system (CNS; Camara, Nathan, & Puente, 2000). This proportion is likely to be even higher for referrals from psychiatric and neurological settings. Information derived from these sorts of assessments might serve as an early warning sign that, if positive, would then result in a more in-depth medical, neuropsychological, or neurological assessment and/or further monitoring of the patient. Examples of the types of situations where screening might be important would be among persons who abuse substances, persons exposed to neurotoxic substances, or elderly populations where the distinction between depression and organically based dementia might be crucial. Additional situations might occur with a school psychologist who is trying to understand why a student is performing poorly, workers' compensation cases in which brain damage might be suspected, or screening for brain damage among psychiatric populations. Each of these situations would require that the assessing clinician be sensitive to the expression of brain impairment, methods of assessing for it, and the patterns of behavioral and test results that would suggest the presence of such impairment. Additionally, more and more neuropsychological and cognitive screening is being used to monitor progress of patients, either improvement based on treatment or decline based on illness, as more data emerge on the links between medical and psychological illnesses and cognitive functioning.

This chapter provides introductory knowledge and strategies for screening for CNS functioning. To that end, the chapter provides introductory knowledge into the area combined with a working knowledge of two major screening instruments, the Bender Visual-Motor Gestalt Test, Second Edition (Bender-2) and the Repeatable Battery for the Assessment of Neuropsychological Status Update (RBANS). Inclusion of this information is based on the premise that, at a minimum, professional psychologists must have as a core competency knowledge on how to screen for CNS complications. If a more in-depth coverage is required, readers are referred to Groth-Marnat's (2000a) *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration*.

When appraising clients with suspected CNS problems, it is important to appreciate that the behavioral manifestation of these problems can be extremely heterogeneous. Some persons with brain damage might have specific signs, such as aphasia, neglect

of a portion of their visual field, or word-finding difficulties. In contrast, others might have widespread impairments, such as a general lowering of cognitive abilities or difficulty regulating their behavior. Expression of deficits might range from being extremely subtle to being quite severe. The practical implication is that many screening tests for neuropsychological impairment are likely to assess for a narrow range of abilities. If a client has deficits outside this range, the test is not sensitive to that particular area of difficulties. The result is a high number of false negatives. Indeed, this problem has plagued most screening devices. For example, a test such as the original Bender was primarily a test of visuoconstructive abilities. Clients with a wide range of other difficulties are likely to perform quite well on the Bender with the resulting danger that the clinician might erroneously conclude they were not organically impaired. The update to the Bender-2 added several additional CNS-related abilities, though it remains primarily a visuospatial test. The RBANS covers a much wider range of abilities and, as such, is more likely to be sensitive to a wider range of CNS complications.

The presence of false negatives (or false positives) depends in part on the “narrowness” versus the “width” of the test. For example, a test that measures a specific function, such as ability to name objects, is quite narrow in its focus. Clients who do poorly on such a test would most likely be experiencing neuropsychological impairment (true positives). However, there are also many persons who, despite having neuropsychological impairment, do quite well on such a test and may be misclassified as normal (false negatives), because neuropsychological impairment may (and often does) leave this skill intact while affecting other domains of functioning. The sign of object naming, thus, is too specific to screen for neuropsychological impairment. If another test is used that casts a wider net by using more general indicators (e.g., concrete thinking, impaired immediate memory, distractibility), not as many persons with neuropsychological impairment will be missed (fewer false negatives). However, many people will be labeled as having brain damage who do not have such damage (many false positives). This is likely to be particularly true for patients with severe psychiatric illnesses. Indeed, neuropsychological tests have had a notoriously difficult time distinguishing individuals with psychosis, especially chronic schizophrenia, from persons with brain damage because they often appear quite similar on test performance (Mittenberg, Hammeke, & Rao, 1989).

The two general strategies in neuropsychological assessment are (1) a qualitative or pathognomonic sign approach and (2) the use of quantitative cutoff scores. The pathognomonic sign approach assumes the existence of distinctive behaviors (signs) indicative of brain damage. Distortions in reproducing designs or reproducing the same design repeatedly (so-called perseverations) are examples of such signs. Additional ones might be aphasias, line tremor, rotating a design, difficulty with serial subtraction, clang responses (e.g., the word *ponder* meaning “to pound”), neglecting a portion of a visual field (visual neglect), or difficulty distinguishing whether a stimulus is on the right or the left when receptors are stimulated at the same time (suppressions on bilateral, simultaneous stimulation). In contrast to the sign approach is the use of cutoff scores, which optimally separates a person’s performance into either a brain-damaged or normal range. The use of cutoff scores is a major feature of the Halstead-Reitan Neuropsychological Test Battery (HRNTB; Broshek & Barth, 2000; Reitan & Wolfson, 1993) and the Neuropsychological Assessment Battery (Stern & White, 2003).

Similar to other psychological tests, moderator variables such as age, education, and premorbid intelligence are related to neuropsychological test performance. Thus, it has been recommended that cutoff scores for determining impairment should use norms corrected for age, education, and sometimes gender. These norms are available in a variety of sources, including the *Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery* (Heaton, Miller, Taylor, & Grant, 2004), *Handbook of Normative Data for Neuropsychological Assessment* (2nd ed.; Mitrushina, Boone, Razani, & D'Elia, 2005), and *Compendium of Neuropsychological Tests* (3rd ed.; E. Strauss, Sherman, & Spreen, 2006).

HISTORY AND DEVELOPMENT

Neuropsychological assessment as a well-defined discipline began in the 1950s with the work of Halstead, Reitan, and Goldstein in the United States; Rey in France; and Luria in the Soviet Union. In the United States, the experimental and statistical orientation of American psychology was reflected in test design and use. Norms were refined and used for comparison with an individual patient's performance. Optimal cutoff scores were developed to distinguish impaired from normal performances. In particular, the HRNTB grew out of an original 27 tests that Ward Halstead selected in the belief that they measured cerebral functioning based on "biological intelligence." Halstead reduced these to 10 tests, and Reitan (1955a) later reduced these to 7. Cutoff scores were developed on these tests, and, based on the proportion of tests in the impaired range, an Impairment Index could be calculated.

Early success was achieved with the HRNTB in distinguishing not only the presence of brain damage but also the location and nature of lesions (Reitan, 1955a). During the days before sophisticated neuroradiological techniques, this information was extremely useful. These efforts emerged into an emphasis on what has sometimes been referred to as the three Ls of neuropsychology: lesion detection, localization, and lateralization. In contrast, the study of diffuse impairment was relatively neglected in favor of the stronger emphasis on focal involvement.

Concurrent with the developments in the United States was the work of Alexander Luria in the Soviet Union and André Rey in France. They relied extensively on close patient observation and in-depth case histories. They were not so much interested in what score a person might have obtained but rather why the individual performed in a certain manner. Their work epitomized the flexible, pathognomonic sign or qualitative approach. Rather than developing a series of quantitatively oriented tests with optimal cutoff scores, Luria emphasized a series of "procedures" that he believed would help the client to express relevant behavioral domains related to CNS impairment. His approach relied far more heavily on clinician expertise and observation than formal psychometric data. Although somewhat controversial (see K. Adams, 1980), these procedures were formalized and standardized into the Luria-Nebraska Neuropsychological Battery (Golden, Purisch, & Hammeke, 1985).

From these early beginnings, two distinct strategies of approaching neuropsychological assessment emerged. One was the comprehensive battery approach pioneered by Halstead and Reitan and formalized into the HRNTB; the other was a more

flexible, qualitative, hypothesis-testing strategy as represented by Goldstein and Luria. Each of these approaches has different strengths and weaknesses (see Bauer, 2000; Jarvis & Barth, 1994; Russell, 2000). The battery approach has the advantages of assessing both strengths and weaknesses for a broad spectrum of behaviors, is easier to use for research, is consequently more extensively normed and researched, can be administered by trained technicians, and is easier for students to learn. Its disadvantages are that it is typically quite time consuming, may overlook the underlying reasons for a client's specific test scores, and is more difficult to tailor toward the unique aspects of the client and referral question. The contrasting qualitative hypothesis-testing approach has the advantages that it can be tailored to the specifics of the client and referral question, emphasizes the processes underlying a client's performance rather than a final score, and is quite time efficient. Measurements of a client's strengths, weaknesses, or certain reasons for ambiguous responses can be pursued in more depth according to decisions made by the examiner. Weaknesses frequently attributed to this approach are that, in practice, it focuses on a client's weaknesses, relies too extensively on clinician expertise, is more difficult to research, is consequently not as extensively researched, and provides a narrower slice of a client's domains of functioning.

Despite the preceding somewhat polarized description, two trends indicate an integration of the quantitative psychometric and the qualitative hypothesis-testing strategies. First, in practice, most neuropsychologists use a combination of the strategies. Surveys of practice indicate that the vast majority of clinical neuropsychological assessments use a "flexible-fixed" battery comprising a relatively short "fixed" or core battery combined with additional flexible tests that can be selected based on the uniqueness of the client and specifics of the referral question (Sweet, Moberg, & Suchy, 2000). The second trend is the development of objective, in-depth, computerized scoring systems that help clinicians understand the underlying qualitative processes a client undergoes in responding to test items (e.g., scoring for the California Verbal Learning Test; Delis, Kramer, Kaplan, & Ober, 2000).

Concurrent with the development of the early testing procedures and batteries, there was also an emphasis on brief screening instruments. The Bender Visual-Motor Gestalt Test was one of the earliest of these. It was first developed by Lauretta Bender in 1938 and included nine designs that a client was requested to reproduce. A similar but more complex visuo-constructive test was devised by Rey in 1941 and expanded by Osterrith in 1944. It has since become refined and referred to as the Rey-Osterrith Complex Figure Test (Meyers & Meyers, 1996; Visser, 1992). Examinees are first asked to complete the copy of the drawing while it is directly in front of them, then they are requested to make a second reproduction of the drawing from memory. Rey also developed the Rey Auditory-Verbal Learning Test (Rey, 1964), which primarily screens for difficulties with short-term auditory memory. Clients are instructed to recall a series of words that are read to them and then repeat back as many of the words as possible. A final example of an early screening test for attentional difficulties is the Stroop procedure (A. R. Jensen & Rohwer, 1966; Stroop, 1935). This test presents clients with a series of written names of colors, but some are written in the same color ink and some in a different color ink from the written name of the color given (see Ponsford, 2000). For example, the word *green* might be written in red ink.

The client is then asked to read the list and give the name of the color of the ink (e.g., red), regardless of the word that is written (e.g., “green”).

A frequent goal of many of the early screening tests was to differentiate between organic and functional (psychological) difficulties. Thus, a referral question was sometimes expressed in terms of “ruling out organicity” or to “differentiate between organic versus functional causes.” More recently, the appropriateness of this goal and the assumptions behind it have been questioned. In particular, the distinction between many functional and organic disorders has gradually disintegrated. For example, early conceptualizations of schizophrenia considered it functional. In contrast, current research supports strong genetic, biochemical, and structural correlates in a substantial proportion of individuals with schizophrenia (Saran, Phansalkar, & Kablinger, 2007). A second factor is that advances in neuroradiological and other neurologically oriented techniques have greatly refined the diagnosis of brain damage. As a result, the use of neuropsychological techniques in diagnosis has been deemphasized. In contrast, referrals from neurologists and psychiatrists are more likely to request information regarding the nature of already identified lesions.

A further change over time has been that, rather than focusing on measurement, there has been greater emphasis on application (Ponsford, 1988; Stringer & Nadolne, 2000). Accordingly, it is no longer sufficient merely to state that a client is experiencing cognitive deficits in certain areas. Instead, answers to more functionally relevant questions are being required, such as the client’s employability, potential for responsiveness to rehabilitation, and the need for certain environmental supports (Sbordone & Long, 1996). Each of these questions can be clarified by considering the differences between impairment and disability. *Impairment* typically reflects normative comparisons and test data. In contrast, the functionally relevant term *disability* more closely takes into account the context of the client, including his or her circumstances, environment, and interests. For example, a client might be statistically in the impaired range on tests requiring sequencing, but if his or her occupation required primarily visuospatial skills, he or she might be able to continue functioning effectively. In contrast, a computer programmer who developed an equal level of sequencing difficulties would be likely to be quite disabled by this problem. Clinicians are increasingly expected to work with both the test data and the specifics of the client to translate the impact of any test-related impairment into a better understanding of the meaning it might have for the client in his or her everyday life. Understanding the meaning of test performance for everyday functioning may also require using methods of analysis other than psychological tests, such as the ratings of relatives, ward observation charts, and simulations (Knight & Godfrey, 1996; Sbordone & Guilmette, 1999).

Consistent with these points is that more recent emphasis has been not so much on measuring “organicity” or “brain damage” but rather on assessing different functions or domains. Possible domains might include attention, short-term memory, or visuoconstructive abilities. Thus, “brain-sensitive” screening tests should not be considered to be tests of brain damage, but rather tests of certain functions that *may* be consistent with CNS problems. Many such instruments have been developed. One representative screening test is a seven-test battery composed of Trail Making, finger-tapping speed, drawing a Greek cross, the Pathognomonic Scale

of the Luria-Nebraska Neuropsychological Battery, the Stroop, and the Logical Memory and Visual Reproduction subtests of the Wechsler Memory Scale (Wysocki & Sweet, 1985). Total administration time for this battery is approximately 60 minutes. Another representative screening system is the Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS; Prigatano, Amin, & Rosenstein, 1992a, 1992b). Its purpose is to determine whether patients are capable of taking other neuropsychological tests. It evaluates their level of self-awareness, provides qualitative information regarding cognitive functioning, and assesses a wide range of cerebral functions. The entire procedure typically takes 10 to 15 minutes to complete. There have also been two abbreviated versions of the Halstead-Reitan Battery by Golden (1976) and by Erickson, Caslyn, and Scheupbach (1978).

In addition to these procedures, several short batteries have been developed for reviewing possible neuropsychological impairment with specific types of disorders. Batteries for the evaluation of neurotoxicity include the California Neuropsychological Screening Battery (Bowler, Thakler, & Becker, 1986), Pittsburgh Occupational Exposure Test (C. Ryan, Morrow, Parklinson, & Branet, 1987), and Individual Neuropsychological Testing for Neurotoxicity Battery (R. M. Singer, 1990). Similar to the previous screening batteries, each of these uses a combination of previously developed tests, such as Trail Making and portions of the Wechsler intelligence scales. Assessment and monitoring of some of the more important domains of dementia might be achieved with the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Battery (Morris et al., 1989) or the Dementia Assessment Battery (Corkin, Growdon, Sullivan, Nissen, & Huff, 1986). A similar specialized battery for detecting the early signs of AIDS-related dementia is the NIMH Core Neuropsychological Battery (Butters et al., 1990).

INTERVIEWING FOR BRAIN IMPAIRMENT

Although tests can be quite useful, the strongest tool for a clinician in assessing brain impairment can often be a clear, thorough, and well-informed history. One of the major factors guiding such a history is understanding the types of behavior that are likely to reflect neuropsychological impairment (see Sbordone, 2000a). Table 12.1 provides a summary of some possible behavior changes indicative of impaired brain processes. The presence of one of these behavioral changes is not sufficient in and of itself to diagnose pathology, but the presence of several would suggest such a process. These behaviors *may* relate to neuropathology, but they may be indicative of other problems as well. An additional tool in extracting the range of possible symptoms is a checklist of potential areas of difficulties that the client can easily complete. Such a checklist might be informally developed by a clinician through simply listing all potentially problematic behaviors, such as difficulties with memory, finding the right word, difficulty organizing thoughts, or confusion. Alternatively, a checklist is commercially available that allows clients to detail the full range of their symptoms (Neuropsychological Symptom Checklist; Schinka, 1983). Any items a client endorses can be further explored in the interview to determine the nature of the symptoms as well as their onset, frequency, intensity, and duration.

Table 12.1 Examples of Behavioral and Emotional Changes That May Indicate Pathological Processes in the Brain

Domain	Behaviors/emotions
Attention	Short-term memory complaints Problems staying focused Difficulty shifting attention Repetitive behaviors (perseveration)
Language	Difficulties with reading, writing, or arithmetic Reversals of numbers or letters Problems understanding spoken or written information Word-finding difficulties Difficulty pronouncing words
Memory ¹	Short- or long-term memory problems Memory acquisition, consolidation, retrieval Memory problems that are auditory/verbal or visual/spatial
Spatial	Poor spatial judgment Poor orientation related to spatial material Difficulty with manual skills (i.e., repairs) Problems distinguishing right from left Not attending to left or right visual field
Executive ²	Difficulty planning Apathy Poor awareness of social impact Difficulty multitasking
Motor	Poor fine motor coordination Tremors Clumsiness Weakness on right or left side of body
Emotional/Behavioral ³	Change in grooming (sloppy, not bathing, overly fastidious) Inappropriate social behavior Change in activity level Change in eating, sex, drinking

¹Note that it is typically difficult to distinguish between memory problems due to neuropathology of the brain and emotional distress caused by environmental factors.

²Poor executive functioning can also reflect the apathy and hopelessness consistent with depression.

³Change in emotional functioning is often difficult to differentiate between neuropathology and external events.

A family history should focus on some of the general areas outlined in Chapter 3. The family history for neurological and/or psychiatric complaints should receive particular attention. A family history that includes conditions with a known or suspected genetic component, such as schizophrenia, early-onset Alzheimer's disease, Huntington's chorea, or hypertension, should alert the clinician that similar processes may be occurring with the client. The presence of any early deaths in the family, learning disabilities, or mental retardation would also be important to consider. Because some types of clients have difficulty recalling detailed information, relevant family members might be contacted to help obtain, elaborate on, or verify information.

Prenatal and early personal history are also important areas for consideration. The client's prenatal environment might have involved relevant events, such as exposure to alcohol, drugs, pesticides, solvents, or dyes. Complications during pregnancy and birth, such as low birthweight, forceps birth, premature birth, or difficulties related to any anesthetics used, should also be considered. Early developmental milestones, including the age at which the client sat upright, walked, and talked, should be noted and ideally verified with an outside source. Academic history is particularly helpful in determining the person's premorbid level of functioning. Favorite and worst subjects, grades obtained, and highest level of education are all significant. Assessing for possible attentional or learning difficulties is also essential. School records often provide useful information, especially when they are objective, and it is always best practice to verify a client's claims related to his or her premorbid level of functioning.

A client's occupational history also helps evaluate his or her premorbid level of cognitive and social functioning. Each occupation requires certain skills that might have implications for interpreting test results. For example, test scores indicating average verbal skills would mean something quite different for an unskilled laborer than for a successful attorney. Average scores might be consistent with the former but could very well reflect impairment for the latter. It might also be relevant to note whether the person's occupation has resulted in exposure to potentially neurotoxic substances, such as organic solvents, insecticides, lead, or mercury. If so, the occupational precautions used and occurrence of any potential exposure incidents would need to be determined. Learning about current and past interests and hobbies can help the clinician develop a more complete portrayal of the person.

The client's medical history and any available medical records should be obtained from the client as well as from relevant persons close to him or her. The central focus of such a review is to attempt to determine whether the current symptoms can be accounted for based on this history. A person might have had a recent head injury, but inferring that his or her symptoms are partially or wholly the result of this injury might be more difficult. The history might include previous head injuries, high fevers, learning disabilities, or exposure to neurotoxic substances. Any history of a head injury should include details as to the last memory the patient had before the injury, recall of the injury itself, the length of time the person was unconscious, and the first memory following the injury. Any behavioral changes (e.g., irritability, poor memory, confusion) should be carefully documented. Further relevant medical complications might include history of high fevers (e.g., $103^{\circ}\text{F}/40^{\circ}\text{C}$) or significant infectious diseases (e.g., meningitis, encephalitis, HIV/AIDS), thyroid dysfunction, diabetes, epilepsy, hypoxia, suicide attempts, hypertension, or neurosurgery for complications

such as tumors or aneurysms. If the patient has undergone any surgery, details should be obtained related to anesthetic use, complications, possible loss of consciousness, psychosocial changes following the surgery, and the nature and duration of these changes. Headaches, especially if accompanied by neuropsychological complaints, might suggest a tumor or a vascular disorder. Drug and alcohol use also needs to be carefully documented along with possible changes in prescription or nonprescription medication. Any current or past psychiatric difficulties might also complicate a client's presentation of neuropsychological symptoms.

Any neuropsychological history should carefully document current complaints and current overall life situation. Each symptom should be described along with its onset, frequency, duration, intensity, and any changes over time. Often much of this information can be accessed by asking the client when the symptom first appeared and how it has changed over time. For example, the abrupt onset of neuropsychological complaints with no clear-cut trauma suggests the possibility of a cerebral vascular accident. In contrast, gradual change might suggest a dementing condition or a slow-growing tumor. Discrete, temporary symptoms suggest transient ischemic attacks. A complicating factor is that clients vary in regard to their awareness of symptoms. Some might be preoccupied with symptoms, others might be indifferent, while still others might be aware of some difficulties but relatively unaware of others. This varying extent of awareness would require that the interviewer refer to medical records and relevant persons in the client's life. Doing so would be especially important in conditions such as dementia or frontal lobe impairment, in which clients might be both unaware of their deficits and inaccurate regarding details of their personal history (desRosiers, 1992; Gilley et al., 1995). A client's sexual functioning can often reveal relevant information related to neuropsychological status. Changes in sexual desire might be related to certain medications, growth of tumors in specific areas, affective disorders, infectious diseases, exposure to neurotoxins, or head injuries (especially with frontal lobe involvement). It is also important for clinicians to investigate the psychosocial factors that might be related to symptoms. Stress, depression, and family turmoil might either cause or serve to exacerbate "neuropsychological" symptoms such as problems with concentration and memory, confusion, and irritability (Burt, Zembar, & Niederehe, 1995; Gorwood, Corruble, Falissard, & Goodwin, 2008; Sherman, Strauss, Slick, & Spellacy, 2000). Finally, legal complications might be intricately entangled with symptoms. This is especially true for cases involving litigation or workers' compensation.

Whereas the preceding suggestions represent a variety of areas that can be explored flexibly, several structured interviews and questionnaire formats are currently available. The Neuropsychological History Questionnaire (Wolfson, 1985) is an easily completed, 37-page, comprehensive series of questions to be answered by the client. It includes topics such as referral information, academic history, medical and general history, and present status compared with pre-injury/pre-illness status. The Neuropsychological Status Examination (Schinka, 1983) includes a similar organization of topics but is a semistructured interview in which most of the questions are asked by the interviewer. The Neuropsychological Status Examination also includes the previously mentioned Neuropsychological Symptom Checklist, which provides a brief self-report of symptoms that can be used to assist the interview. An extremely detailed and long (3–5 hour) structured questionnaire is the Neurobehavioral Assessment Format

(A. Siegel, Schechter, & Diamond, 1996). Additional useful tools might be brief, simple rating forms such as the Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975), Neurobehavioral Rating Scale (Levin et al., 1987), or Patient Competency Rating (Prigatano, 1986). Any of these structured formats requires an examiner to integrate the information into the unique characteristics of the client and relevant test data.

DOMAINS OF NEUROPSYCHOLOGICAL FUNCTIONING

Neuropsychological tests have traditionally been organized according to five domains: attention, language, memory, visuospatial, and executive functioning. Sometimes a measure of achievement is also used and included as part of language functions. These domains cover the primary aspects of a person’s cognitive functioning. Table 12.2 lists these domains along with measures that are frequently used to assess them. As can be seen, many of the subtests of the Wechsler intelligence scales are quite useful in providing information for these domains. Often professional psychologists first administer the Wechsler intelligence scales and then amplify them with more specific cognitive tests. For example, the Wechsler Memory Scale—IV might be used to obtain more in-depth information related to memory functions. Similarly, the Bender–2 might be administered to better understand a client’s visuospatial abilities. Relevant background information for conducting neuropsychological screening involves a brief understanding of these domains.

Attention

The maintenance of an optimum amount of mental activity involves a complex variety of functions related to filtering, selecting, focusing, shifting, tracking, and sustaining (see Baddeley, 2003; Ponsford, 2000). Because there is typically a large amount of

Table 12.2 Domains of Cognitive Functioning and Frequently Used Tests to Measure Them

Domain	Tests
Attention	Arithmetic, Digit Span, Letter-Number Sequencing, Cancellation, Stroop Color Word Test, Conners Continuous Performance Test—3 (CPT–3)
Language	Vocabulary, Comprehension, Information, Aphasia Screening Test, Boston Naming, Controlled Oral Word Association Test, Peabody Picture Vocabulary Test—IV, RBANS
Memory	Wechsler Memory Scale, Rey Auditory Verbal Learning Test, Bender–2 (recall), Rey-Osterrieth (recall), RBANS
Visuospatial	Block Design, Picture Concepts, Matrix Reasoning, Picture Completion, Bender–2 (perception and copy), Rey-Osterrieth Complex Figure Test, Judgment of Line Orientation, RBANS
Executive Functioning	Interview/history, Delis-Kaplan Executive Function System, Category Test, Wisconsin Card Sorting Test, Trail Making Test, Behavioral Assessment of the Dysexecutive Syndrome

available information to attend to, a person must be able to filter this potential information and attend to only the most relevant sources. Any irrelevant information must be ignored. This filtering, selecting, and focusing process is still not sufficient in and of itself. The abilities to sustain and shift attention also play a large role in functioning. Attention must strike a balance and be neither overly focused nor too ready to shift. An individual who becomes too focused expresses this symptomatically in perseverations. Such persons then experience difficulty shifting their attention to a new task and are therefore likely to continue with a behavior beyond the point at which it is useful. Conversely, people who shift their focus too readily express this symptomatically in distractibility.

Because of the complexity and interrelationship with other tasks, attention is quite sensitive to the effects of CNS complications. It is thus one of the most frequently reported disturbances associated with cerebral impairment (Lezak, 1989). The most basic form of assessment for attentional deficits is through simple reaction time tasks. For example, reaction time has been found to be sensitive to the effects of head trauma (Van Zomeren & Brouwer, 1990), solvent exposure (Groth-Marnat, 1993), and the early impact of dementia (Teng, Chui, & Saperia, 1990). As attentional tasks become more complex, they become progressively more sensitive to the impact of neuropsychological dysfunction.

A good starting place to understand a client's attentional abilities is to look at the WAIS-IV/WISC-V Working Memory Indexes and the subtests comprising these indexes (Digit Span, Letter-Number Sequencing, Cancellation); it is important to note that while these tasks require more than just attentional ability, impaired attention will necessarily affect performance on them. The Stroop is also a frequently used measure of attention (Jensen & Rohwer, 1966; Stroop, 1935). Clients are presented with a series of names of colors written either in the same or different color ink from the written name of the color given (see Ponsford, 2000). Clients then are asked to read the list and give the name of the color of the ink (e.g., red) rather than reading the word (e.g., "green"). In order to do well, they must be able to disengage their attention from the "pull" of the word in order to name the color of the ink.

The Conners' Continuous Performance Test, Third Edition (CPT-3; Conners, 2015) is one of the most widely used instruments to assess attention. It is a computer-administered test that requires examinees to respond to targets and nontargets that flash on the screen. Specifically, examinees are told to press the spacebar or mouse button when they see any letter flash on the screen other than "X" and to do nothing whenever they see an "X" flash on the screen. The task continues for 14 minutes, with different configurations and intervals of time between flashing letters. Performance on this test reflects ability to attend, sustain that attention during a relatively uninteresting task, remain vigilant, and control impulses. As such, it is a good measure of attention, concentration (sustained attention and vigilance), and impulsivity.

Language

Disturbances of verbal functions are frequently associated with brain damage, particularly when the damage is to the left hemisphere. As a result, any review of neuropsychological functions needs to assess verbal abilities. Often this involves

assessing the academic skills that are frequently associated with verbal abilities. The most common disturbances are the aphasias (impaired speech, writing, or understanding spoken or written language) and problems with speech production. These disorders can involve extremely diverse difficulties, including poor articulation, loss of verbal fluency, word-finding difficulty, poor repetition of words or sentences, loss of grammar and syntax, misspoken words (paraphasias), poor auditory comprehension, reading difficulties, and impaired writing (Goodglass & Kaplan, 1983).

Due to the variety of these symptoms, neuropsychological screening can assess only a relatively small number of them. For a full assessment of aphasic and related disorders, several comprehensive batteries are available, including the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983), Communicative Abilities in Daily Living (Holland, 1980), and Multilingual Aphasia Examination (Benton & Hamsher, 1989). In contrast to these formal, comprehensive batteries, Lezak, Howieson, Bigler, & Tranel (2012) recommended an informal and general clinical review of six major functions:

1. *Spontaneous speech*. Observe how clients initiate, articulate, and organize their speech.
2. *Speech repetition*. Ask clients to repeat words, phrases, and sentences. In particular, this might include repeating difficult words, such as *Massachusetts* or *Methodist Episcopal* (see Reitan & Wolfson's, 1993, Aphasia Screening Test), to assess for disorders of articulation.
3. *Speech comprehension*. Request that clients answer simple questions (e.g., Is a ball square?) or obey simple commands (e.g., point to specific objects, put their hands on their chins).
4. *Naming*. Ask clients to name common objects, colors, letters, and actions.
5. *Reading*. Have clients read aloud; for comprehension, have them explain what they have read.
6. *Writing*. Request that the client copy, write to dictation, and compose a sentence.

The relative difficulty of the verbal tasks should be tailored to additional information regarding client symptoms and behaviors. For example, it would be neither necessary nor appropriate to request that a client with merely mild deficits obey quite simple commands or name common objects. Useful information regarding verbal abilities can often be derived from relevant Wechsler intelligence subtests (Vocabulary, Similarities, Information, Comprehension). As the expressive vocabulary task that is measured by the Wechsler Vocabulary subtest is extremely useful in detecting multiple verbal problems, it is important to note that low scores could be due to multiple problems, including poor comprehension and knowledge or difficulty expressing oneself clearly. A good way to understand low scores on Vocabulary is by administering an additional test of receptive vocabulary, such as the Peabody Picture Vocabulary Test, Fourth Edition (PPVT–IV, Dunn & Dunn, 2007), which assesses word knowledge without the added ability of clear expression. The PPVT–IV presents four pictures to the examinee, the clinician says a word, and the examinee must respond with which picture best represents the word. When scores are low, it is a clearer indication of difficulty with

words and understanding language than a low Vocabulary score alone. When scores are adequate and Vocabulary scores are low, the problem is more likely one with expressive language, such as clearly verbalizing ideas, than with understanding the words themselves.

The most frequently used educational achievement battery in clinical neuropsychology (Camara et al., 2000) is the Wide Range Achievement Test, Fourth Edition (WRAT-4; Wilkinson & Robertson, 2007). The battery is easy to administer, covers a wide range of ages (12–75 years), and provides scores for spelling, reading, and arithmetic. Each score can be conveniently portrayed as school grade equivalents, standard scores, or percentiles. However, the battery assesses a somewhat narrow range of abilities in these domains and thus should be used only as a general screening instrument. An increasingly popular, more in-depth assessment of achievement can be obtained from the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Psychological Corporation, 2009), which measures reading, writing, mathematics, and oral language comprehension, including following directions. Additionally, the Woodcock-Johnson Tests of Achievement, Fourth Edition (Woodcock-Johnson IV; Schrank, Mather, & McGrew, 2014) offers a more comprehensive battery of achievement tests that can be tailored to a specific referral question.

Memory

The types and procedures of memory and learning are complex (see Baddeley, 2003; Baddeley, Kopelman, & Wilson, 2002; Helmes, 2000). Aspects of these processes might include sensory memory, short-term memory, rehearsal, long-term memory, consolidation, recall, recognition, and forgetting. In addition, memory and learning can be divided into two major subsystems: declarative memory, which refers to learning about information, objects, and events; and procedural or implicit memory, which refers to automatic, habitual responses. Each of these subdivisions has somewhat different anatomical structures. Additional useful subdivisions of memory are verbal versus spatial, automatic versus effortful, and semantic versus episodic. Studies of patients with brain lesions indicate that memory can be further divided into extremely specific subareas based on functions such as sensory modality (verbal, tactile, auditory, etc.), type of material (verbal, motor skill, etc.), and content of information (numbers, letters, pictures, names, faces, etc.; see Baddeley, 2003; Baddeley et al., 2002). For the practitioner, providing a truly comprehensive evaluation of memory functions is a daunting task.

Fortunately, usually a more limited number of memory domains can provide practitioners with an overview of the general intactness of memory. These include: (1) the extent to which the subject can acquire and retain new material; (2) how quickly material is forgotten; (3) the extent to which competing information interferes with learning; (4) the degree of specificity or generality of the deficit; and (5) the stability or fluctuation of the deficits over time (Walsh, 1994). Ideally, these domains should include measurements of both visual and auditory/verbal material.

One important distinction is between attention versus memory and learning. In some ways, this distinction is inappropriate because attention is a prerequisite for learning to occur. A person who is easily distracted does not effectively learn and remember relevant information or events, as the information may not even enter the person's

awareness. Attention is, therefore, closely linked to learning. However, in other ways attention and learning do represent distinct functions. In particular, it is important to distinguish whether a person is capable of learning but is easily distracted, or whether, even under circumstances in which the person fully attends to a task, he or she still cannot learn effectively. Sometimes clients state that they have a memory problem, but, despite their symptom description, they perform learning and memory tasks quite well under the ideal circumstances that often characterize assessment procedures. In contrast, in real-world situations, they frequently need to exclude a number of distractions and carry on two or more activities simultaneously. Under these conditions, they might have distinct difficulties dividing their attention and, therefore, might not be able to learn and remember particularly effectively. Interviewing them regarding situations in which they do versus do not remember effectively might help the practitioner understand this issue better. In addition, their test performances would be expected to be lower on tasks that load more heavily on attention (Arithmetic, Digit Span, serial 7s or serial 3s) than those that are more pure tests of learning (Wechsler Memory Scales, repeating paragraphs/stories, Bender–2 recall, RBANS).

A good beginning place to assess memory is in the interview. Details regarding basic information, such as personal, family, educational, and employment history, can be pursued. Interviewers might request dates when the client began or finished employment or education, parents' or children's birthdays, or details related to medical history. Some of this information might be compared with more objective sources to determine its accuracy. In addition, behavioral observations, such as pauses, expressions of uncertainty, or confusion, might suggest difficulties with retrieval.

Current research consistently indicates that there is a mild to moderate relationship between memory impairment and depression. An extensive meta-analysis by Burt, Zembar, and Niederehe (1995) found that memory impairment was associated most clearly with inpatients (versus outpatients) and patients with mixed bipolar and unipolar disorders (versus those with purely unipolar disorders). Similarly, Gorwood et al. (2008) found that frequent, long, and chronic states of depression clearly impaired brain functioning in those areas responsible for memory. In addition, negative affective information was more likely to be remembered accurately than material with a positive or neutral emotional tone (Burt et al., 1995). However, memory impairments were also present among populations of adults with schizophrenia and mixed groups of psychiatric patients, but not among patients diagnosed with either anxiety disorders or substance abuse. Interestingly, the association between memory and depression was stronger among younger than older persons (Burt et al., 1995; Gorwood et al., 2008). This is probably because early-onset depression is likely to be more severe and younger persons have a greater amount of memory to lose (higher "ceiling" and lower "floor") than older persons (narrower range between ceiling and floor). More recent research has suggested that although depression does seem to correlate with memory impairment (Norman, Tröster, Fields, & Brooks, 2002), these memory impairments more likely represent problems in attention and consolidation from short-term to long-term memory, rather than memory impairment itself (Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010). It is important to note that the link between memory impairment (and other forms of neuropsychological functioning) and depression is typically of quite a small magnitude (Burt et al., 1995; Sherman

et al., 2000). For example, dementia typically accounts for a far larger proportion of the variance in neuropsychological functioning than depression.

To more fully assess the complex and multifactorial structure of learning and memory, a number of relatively comprehensive memory batteries have been developed. Among the oldest, and certainly the most frequently used (Camara et al., 2000; Rabin, Barr, & Burton, 2005), is the Wechsler Memory Scale (WMS; Wechsler, 1945, 1974; see Chapter 6). Its most recent revision is the Wechsler Memory Scale—IV (WMS-IV; Pearson, 2009c), which has five primary indexes (but only four indexes for the Older Adult Battery) and assesses both verbal and visuospatial functions and includes a delayed recall component (see Chapter 6). Additional relatively comprehensive batteries include the Rivermead Behavioral Memory Test (B. A. Wilson, Cockburn, & Baddeley, 2003), Memory Assessment Scales (J. M. Williams, 1991), and Wide Range Assessment of Memory and Learning—II (WRAML—2; Sheslow & Adams, 2003).

In addition to these comprehensive batteries are a number of brief, narrow, specific tests that measure memory. The Rey Auditory Verbal Learning Test is a relatively short, well-researched, frequently used, individually administered test that presents clients with a series of word lists. Their performance is used to assess short-term verbal memory, the ability to learn new material, the extent to which interference disrupts learning, and the ability to recognize information that might have been previously learned. As the name suggests, however, it is verbally oriented. To include at least some visuospatial memory assessment, the recall administration of the Bender–2 can be used. Clinicians might also consider the Benton Visual Motor Retention Test (Benton, 1974) or the Rey-Osterrieth Complex Figure Test (Meyers & Meyers, 1996; Osterrieth, 1944; Rey, 1941, 1964; Strauss et al., 2006). In addition, the WAIS-IV/WISC-V subtests of Digit Symbol-Coding (incidental learning), Information, Digit Span, and Letter-Number Sequencing include potentially valuable information related to learning and memory. However, it should be stressed that Digit Span and Letter-Number Sequencing are primarily attentional tasks rather than pure learning tests.

Visuospatial Functions

The accurate construction of objects involves intact visual perception, effective visuospatial and visuomotor abilities, and the integration of these skills. Each one of these three areas (perceptual, spatial, motor) might have disturbances that could make visual construction more difficult, and impairment may reside in their integration. Benton (1979) enumerated the primary disturbances as:

1. *Visuoperceptual disturbances.* Impaired discrimination of complex stimuli, visual recognition, color recognition, figure-ground differentiation, visual integration.
2. *Visuospatial disturbances.* Impaired localization of points in space, topographic orientation, neglect of part of a person's visual field, difficulties with direction and distance.
3. *Visuomotor disturbances.* Defective eye movement, assembling, graphomotor performance.

For some patients, these disturbances might occur together, whereas with others, they might occur separately. A patient might have excellent visuoperceptual abilities

but still have significant problems making accurate constructions. At other times, poor perception would lead to or occur in combination with poor constructional abilities. In addition, the ability to draw and assemble objects can be quite variable for a particular patient whose ability to assemble objects might be intact (as in Block Design), but whose drawings of human figures or simpler designs might be quite poor. Finally, some individuals may have unimpaired perceptual, spatial, and motor abilities, but the higher-order ability to integrate these three effectively may be impaired.

Each one of the three primary disturbances is also likely to have somewhat different neuroanatomical pathways. The practical implication is that the clinician should not make any inferences regarding localization of lesion merely by considering a person's overall score on a particular visuoconstructive test. As overall scores are of limited use, important information and the implications for localization can be derived more appropriately from a careful observation of how the client approaches the task and the types of errors the person makes. In general, patients with lesions in their right hemispheres tend to approach visuoconstructive tasks in a fragmented, piecemeal fashion in which they often lose the overall gestalt of the design. In contrast, patients with left-hemisphere lesions are likely to duplicate the overall gestalt of the design correctly but often omit important details of the drawing.

A number of assessments of visuoconstructive abilities are available. The Bender–2 is a simple, straightforward task. Its predecessor, the Bender Visual-Motor Gestalt Test, was extensively researched and frequently used in clinical practice (Brannigan & Decker, 2003; Camara et al., 2000; Lacks, 1999, 2000). The Bender–2 might be further supplemented with a free drawing task, such as a Human Figure Drawing or House-Tree-Person. Other somewhat simpler drawing tasks, such as drawing a clock, bicycle, or Greek cross, have been used frequently. Whereas these tasks involve drawing, the Block Design task requires assembling (rather than drawing) designs with the added factor of a time limit.

Executive Functions

Executive functions involve a person's ability to effectively regulate and self-direct behavior. These functions can be subdivided into volition, planning, purposive action, and effective performance (Lezak et al., 2012; Sbordone, 2000b). For example, patients experiencing significant executive impairments might exist in a semivegetative state in which they rarely initiate much activity, even if other cognitive abilities might be quite intact. Other patients with executive difficulty may have little awareness of their impact on others and thus are unable to effectively direct or regulate their social behavior. Although frontal lobe damage is most typically implicated with executive deficits, damage to subcortical, especially thalamic, regions or the more diffuse damage caused by anoxia or organic solvents can also produce executive impairment (see Sbordone, 2000b).

Despite the importance of executive abilities, they can be overlooked during formal psychological assessment, in part because executive functions can be impaired even when other cognitive functions appear quite intact. As a result, a clinician might look at cognitive test scores, such as a composite IQ, and erroneously conclude that a patient has made a good or even full recovery. There are even anecdotal reports of patients'

IQs actually increasing after frontal lobe damage, although they became quite disabled because of a loss of executive abilities.

Another reason executive functioning might be overlooked is that most formal assessment occurs in a structured situation in which the examiner directs the patient to do certain activities. As a result, the patient's ability to self-initiate might be overlooked. This can be remedied by examiners "structuring" an unstructured situation in which patients can demonstrate the extent, style, and manner in which they would initiate, develop, plan, and monitor their own behavior. A final assessment issue is that, frequently, depression can produce some of the same behaviors (e.g., apathy, flat affect, lack of direction) that occur with executive loss stemming from brain damage. A clinician might, therefore, erroneously conclude that executive dysfunctions were the result of depression rather than brain damage (or vice versa).

Interview assessment of executive function might focus on a patient's articulation of future goals, along with descriptions of recreational activities. Typically, patients with executive difficulties provide little detail about these areas. If they do provide detail, it may be primarily based on reciting their goals and activities before the injury. For this reason, interviewers need to establish what patients' current activities and goals are and, in particular, what they have done recently to pursue these goals. Interviewers might also establish the extent to which patients can realistically pursue these goals, anticipate and plan relevant activities, develop alternative plans, and give direction to actually putting these plans into action. Because poor executive functions are frequently accompanied by lack of awareness, it might be essential to interview family members who have had a chance to observe the patient on a daily basis. Thus, the client's descriptions can be compared with more objective external descriptions.

In the actual examination itself, various types of behavior can provide information. Does the patient initiate and direct any activity, or does he or she tend to be relatively passive? Are there unusual social behaviors (e.g., poor grooming, discussion of irrelevant tangents, inappropriate jokes) that suggest poor awareness of his or her social impact? The examiner must determine whether such behaviors developed postinjury or were premorbid characteristics. Planning abilities might be estimated based on how well such patients organize their human figure drawings, blocks on Block Design, Bender-2 drawings, or stories on the Thematic Apperception Test (TAT). Perseverations suggest poor mental flexibility and difficulty monitoring behavior; the patient may make too many dots on the Bender-2 or find it difficult to understand changes in test stimuli (e.g., slow to understand the requirements of WAIS-IV/WISC-V subtests). Because poor executive functions also include difficulty attending to stimuli while simultaneously performing other tasks, low scores on the Wechsler Working Memory Index or Bannatyne's Sequencing might also reflect poor executive abilities.

A number of informal clinical tests might also help to determine possible executive impairments. For example, the patient might be asked to continue the pattern of a drawing that has various repetitive but alternating small shapes (three circles, two squares, one triangle) and then to repeat this sequence several times (see Goldberg & Bilder, 1987). A similar chain-of-command-type test is having the patient tap the desk with his or her fist, then tap it with the palm, then repeat this pattern several times. A slightly more complicated task might be as follows: The examiner taps his or her foot once, then the patient taps a foot twice. Alternatively, the examiner may

tap a foot twice, then the patient is instructed to tap once (see Lezak et al., 2012). None of these procedures has formal scoring; instead, the examiner must determine, based on observation, whether the patient had relative difficulty with all or any of the activities. Although no single strategy in this section is sufficient to identify executive impairments, collectively, the strategies will help to ensure that this critical domain of functioning is included in a client's assessment.

The Trail Making Test (Army Individual Test, 1944; Reitan & Wolfson, 1993) is a frequently used measure of executive functioning. It is an easily administered, widely used test that requires clients to draw lines connecting consecutively numbered circles (Trails A) followed by a similar task in which they draw lines connecting alternating numbered and lettered circles (Trails B; see Figures 12.1 and 12.2). Shifting mental strategy from Trails A (simple consecutive number responding) to Trails B (alternating sequential numbers and letters) requires a shift in strategy, inhibition of the first, logical strategy (connecting consecutive numbers), and the ability to hold two sequences in mind at once (numbers and letters). Scores are based on the total time it takes to complete Trails A and the total time it takes to complete Trails B.

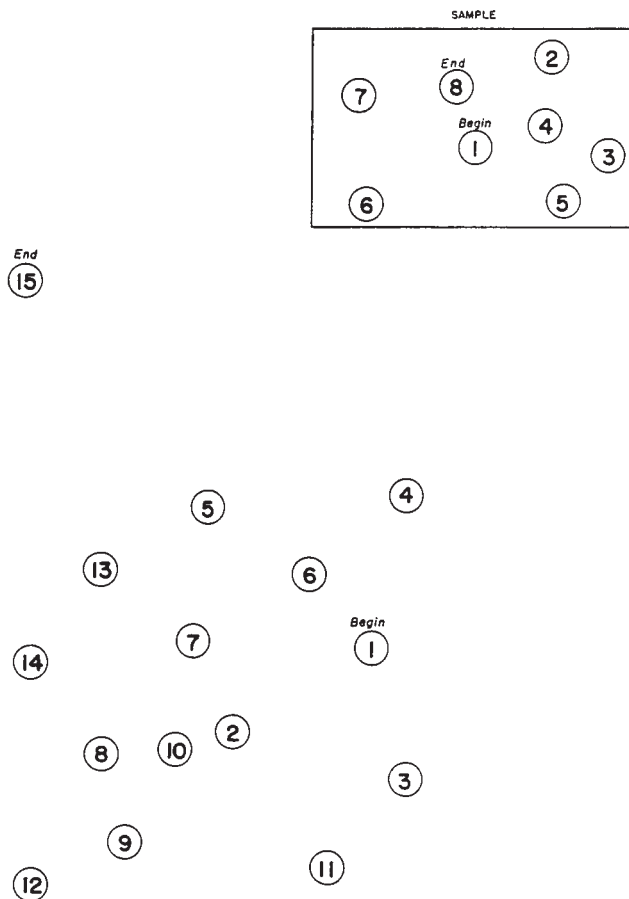


Figure 12.1 Trail Making Part A (abbreviated/child version)

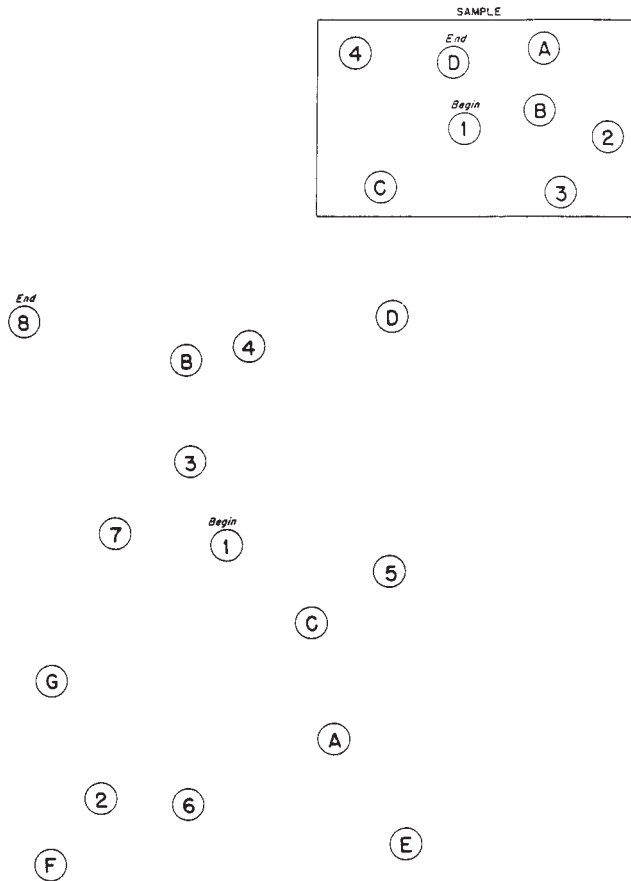


Figure 12.2 Trail Making Part B (abbreviated/child version)

There are also a number of formal comprehensive batteries to assess for executive function. These include the Behavioral Assessment of the Dysexecutive Syndrome (B. Wilson, Alderman, Burgess, Emslie, & Evans, 1999), Frontal Lobe Personality Scale (Grace, Stout, & Malloy, 1999), and the widely used Delis-Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001).

BENDER VISUAL-MOTOR GESTALT TEST, SECOND EDITION

History and Development

The Bender Visual-Motor Gestalt Test (Bender, 1938), usually referred to as the Bender-Gestalt or simply the Bender, has been used extensively as a screening device for neuropsychological impairment by assessing a client's visuoconstructive abilities. The original version consisted of nine designs that were sequentially presented to subjects with the request that they reproduce them on a blank, 8.5-by-11-inch sheet of paper. The examinee's designs were then rated on their relative degree of accuracy

and overall integration. The test's popularity is partly due to the fact that it is brief, economical, flexible, nonthreatening, nonverbal, and extensively researched.

Despite sometimes-equivocal reviews and ambiguous research findings, the Bender has consistently been one of the five or six most frequently used tests (Camara et al., 2000; Kamphaus, Petoskey, & Rowe, 2000). This is consistent with other studies on test usage dating back to 1969. In contrast, Camara et al. found that it was ranked as the 25th most frequently used test by specialty neuropsychologists. This finding likely results from the greater number of options among specialty neuropsychological tests, in addition to the fact that the Bender is not highly regarded among this subgroup.

A wide number of scoring systems for adults and children have been developed for the Bender, each having various advantages and disadvantages. One of the earliest and most widely accepted scoring systems for adults was developed by Pascal and Suttell (1951). Although this system is widely cited in research studies, it has not gained wide acceptance in clinical settings, primarily because of its complexity and time inefficiency. Another early adult system was developed by Hutt in the 1940s and later formally published in 1960 (Hutt & Briskin, 1960). Although his interest in the Bender was primarily for projective personality assessment, he also listed "12 essential discriminators of intracranial damage" (Fragmentation, Closure Difficulty, etc.). Lacks (1984, 2000) adapted this system and provided a detailed scoring manual along with substantial empirical support. In contrast to the Pascal and Suttell (1951) system, it is straightforward and time efficient, typically taking 3 minutes or less to score. Studies using Lacks' system have reported good discrimination between populations with and without brain damage (Lacks, 1984, 1999, 2000; Lacks & Newport, 1980). The system is limited to persons 12 years of age or older.

A system for children was developed by Koppitz (1963, 1975), revised and more recently known as the Koppitz-2 (C. R. Reynolds, 2007). Koppitz carried out an extensive standardization of 1,104 children from kindergarten through fourth grade. Her system provided measures of both developmental maturation and neuropsychological impairment. She cautioned that, for a diagnosis of brain damage, the examiner needs not only to consider the child's scores but also to observe the time required to complete the test, the amount of space used, the child's behavior, and the relative degree of awareness about his or her errors. The original Koppitz system was developed for relatively young children, primarily because the scores of children over the age of 10 were found to no longer correlate with either intelligence test results or age. In addition, after the age of 10, most individuals obtain nearly perfect scores. However, research has indicated that the Koppitz system can be used for adolescents between the ages of 12 and 18, although the relation with age is not nearly as strong as with younger children (McIntosh, Belter, Saylor, Finch, & Edwards, 1988).

The original Bender was not only used to screen for neuropsychological impairment, but many systems for interpreting Bender performance were developed to represent personality characteristics. It is important to note that the research does not support its use for this purpose (e.g., see Holmes, Dungan, & Medlin, 1984; Holmes & Stephens, 1984; Sattler, 1985, 2014). Global ratings for specific characteristics, which typically sum a series of indicators (e.g., size increases, collisions, scribbling), have had greater support in the literature. For example, accurate discriminations have been made

for impulsivity by comparing total scores for impulsive versus nonimpulsive indicators (Oas, 1984). Likewise, Koppitz (1975) listed emotional indicators that have been found to be good predictors of the general presence of psychopathology when three or more are present (Koppitz, 1975; Rossini & Kaspar, 1987). However, this system was not effective at discriminating between individuals with obsessive-compulsive disorder and those without (Kohli, Rana, Gupta, & Kulhara, 2015). Thus, the Bender has generally been found to be valid in predicting the absence or presence of psychopathology based on clusters of indicators, but it has not been as successful for specific, single indicators. With the possible exception of impulsivity and anxiety, the Bender is generally ineffective in identifying specific personality characteristics or specific psychiatric diagnoses.

The Bender was significantly revised and published in 2003 as the Bender Visual-Motor Gestalt Test, Second Edition (Bender-2; Brannigan & Decker, 2003, 2006). This revision was partially based on the earlier work of Brannigan and Brunner (1989), who used a scoring system that rated the quality of each of the reproductions of the designs. The ratings were then totaled to form an overall or global rating of all of the designs. This global rating system is in contrast to other systems (e.g., Hutt, Lacks, Koppitz) that focused on types and numbers of specific errors (e.g., perseverations, distortions, rotations). The 2003 revision retained all of the original designs. However, four easier items were added to create a lower “floor,” and three more difficult items were added to create a higher “ceiling.” These additions enabled the test to be normed on and used with young children down to the age of 4 and adults up to the age of 85+. Children below the age of 8 are given the four easy designs plus the original nine designs (items 1–13). Children above the age of 8, along with adolescents and adults, are given the nine original designs plus the three more difficult ones (items 5–16). The Bender-2 begins with a copy phase, in which the client is shown the designs and requested to copy them. A recall phase follows, in which the client is requested to draw as many of the designs as possible from memory. Two brief, additional subtests were added to assess pure perceptual functioning and pure motor functioning. On the former, examinees are asked which of four choices looks identical to a simple design presented. This assesses perceptual ability without the integration of visuomotor functioning with perception, which is required on the copy and recall tasks. The motor subtest asks examinees to use a pencil to connect dots, without lifting the pencil and without touching the line boundaries that surround the dots on increasingly narrow, and thus more difficult, designs. The subtest basically only requires visuomotor ability, as it requires fine motor control and the ability to connect dots in a very planned, specific way. Adding these two subtests allows the Bender-2 to evaluate where in the process an individual may be struggling—in the perceptual realm, in motor functioning, or in the integration of visuoperceptual and motor functioning (as well as short-term memory, assessed by the recall phase).

Norms for the Bender-2 were based on a large, nationally representative sample that closely paralleled the 2000 U.S. census (based on age, sex, race/ethnicity/geographic region, and socioeconomic level). These included a total of 4,000 participants between the ages of 4 and 85+. In addition, data were gathered from a sample of persons with mental retardation, learning disabilities, attention-deficit/hyperactivity disorder

(ADHD), serious emotional disturbance, autism, Alzheimer's disease, and giftedness. Scores are compared with age-related peers and transformed to standard scores with a mean (*M*) of 100 and a standard deviation (*SD*) of 15.

Reliability and Validity

Reliabilities for the Bender–2 indicate it provides stable, consistent measures (Brannigan & Decker, 2003). Test-retest reliability over a 2- to 3-week interval was .85 (range = .80–.88) for the copy phase and .83 (range = .80–.86) for the recall phase. Split-half procedures indicated the overall validity was .91 and the standard error of measurement (SEM) was 4.55. The developers also provided evidence that there is moderate to high interrater consistency for trained scorers. Rater agreement for the copy phase was .90 (range = .83–.94) and .96 (range = .94–.97) for the recall phase. Even inexperienced scorers had quite high agreement (Copy phase = .85, Recall phase = .92).

Many studies on the previous version of the Bender indicated it was able to discriminate populations with brain damage from those without brain damage (Hain, 1964; Lacks, 1984, 1999, 2000; Marley, 1982). Studies using the Lacks adaptation have reported diagnostic accuracies of from 64% to 84% with a mean of 80% (Lacks, 1999, 2000; Lacks & Newport, 1980). Its diagnostic accuracy has been questioned, however, when used to assess subtle neuropsychological deficits, such as among many persons with epilepsy, or when a differentiation is attempted between patients with functional psychosis and patients with brain damage (Hellkamp & Hogan, 1985). However, this distinction may be less meaningful as schizophrenia is being progressively conceptualized as an organically based disorder. Further studies have found that Bender performance has been able to differentiate patients with Alzheimer's disease from controls as well as reflect the progression of the disease (Storandt, Botwinick, & Danzinger, 1986). Similarly, Bender scores were able to predict the extent to which patients with head trauma could function independently (Acker & Davis, 1989). While uses for specific purposes have been found and may be promising, the Bender is currently better used as an indicator of gross neurological damage.

Research on the Bender–2 indicated moderate correlations with similar measures, thereby supporting its construct validity. For example, the Bender–2 has shown moderate correlation (between .48 and .65) with the Beery-Buktenica Developmental Test of Visual Organization—IV (Brannigan & Decker, 2003; Volker et al., 2010) and a similarly moderate to high correlation (.75) with the WISC-III Perceptual Organization factor (Decker, Allen, & Choca, 2006). Brannigan and Decker (2003) also reported a .80 correlation with the Koppitz Developmental Scoring System. Moderate correlations have also been found with measures of academic skills including the Woodcock-Johnson reading cluster (.53) and the Wechsler Individual Achievement Test—II (.41; Brannigan & Decker, 2003). Correlations with the WAIS-III Performance IQ was .52 and Verbal IQ was .47 (Brannigan & Decker, 2003). As expected, correlations were higher for nonverbal abilities than for more verbally oriented skills.

Scores on the Bender–2 rapidly increase with age, especially between the ages of 5 and 10 (Brannigan & Decker, 2003; Decker, 2007). They increase more gradually between ages 10 and 15 and then remain fairly stable throughout most of adult life.

Table 12.3 Bender–2 Performance among Selected Clinical Populations

Patient type	Bender–2 <i>M</i> (<i>SD</i>)	Matched sample <i>M</i> (<i>SD</i>)
Intellectually gifted	copy 110.62 (11.35)	102.83 (11.38)
	recall 114.94 (17.10)	102.77 (12.70)
Reading disabilities	copy 92.33 (11.04)	103.51 (11.22)
	recall 92.96 (10.93)	103.57 (12.79)
Alzheimer's	copy 100.53 (9.59)	100.95 (7.86)
	recall 81.10 (7.10)	100.21 (7.36)
ADHD	copy 91.15 (13.34)	104.65 (9.30)
	recall 93.39 (13.66)	103.89 (12.89)
Mental retardation	copy 75.77 (14.14)	103.01 (11.31)
	recall 83.92 (10.04)	101.97 (11.89)

Scores are transformed to a standard score ($M = 100$, $SD = 15$).

All score differences with the matched sample were significant ($p < .001$).

Source: Data derived from Brannigan and Decker (2003).

They gradually decrease between ages 40 and 69 and drop off rapidly from 70 to 80+. This finding suggests performance on the Bender–2 can be used to track developmental changes in visuomotor processes in the 4- to 15-year-old range and again after age 70.

Various groups achieve scores on the Bender–2 that roughly correspond to the severity and nature of their cognitive difficulties (see Brannigan & Decker, 2003). A review of Table 12.3 indicates that intellectually gifted persons have the highest scores and those with mental retardation the lowest. Note that the scores for patients with Alzheimer's disease have copy scores in the average range, indicating that visuoconstructive abilities are relatively spared in the early to middle phases of the disease. As expected, given the memory impairment among these patients, the recall phase was found to be quite low. Research indicates that the Bender–2 has moderate to strong validity.

Assets and Limitations

The Bender–2 is a brief, user-friendly screening instrument that is easily administered and measures visuoconstructive abilities. The original Bender has a long history of clinical use and research support. Given that the nine original designs were retained for the Bender–2, it is likely that much of the early research on the original test can be cautiously transferred to the newer version. The research on the Bender–2 also provides good support for its validity. Data presented in the manual indicate that it can be reliably scored by both experienced and inexperienced examiners. In addition, it has good internal consistency and test-retest reliability.

The Bender–2 has a number of features that suggest an improvement on the earlier version. The norms cover a wide range of ages and are representative of the 2000 U.S. census. Since they include both children and adults, practitioners do not have to use separate scoring systems for children and adults, as was the case with the original Bender. The lower floor and higher ceiling created by including additional easy and difficult items increases the Bender–2's sensitivity to both developmental delays and

acquired cognitive impairment. A final asset of the Bender–2 is the inclusion of norms and formal administration and scoring procedures for the recall phase.

Although the Bender has a good track record of achievements, a number of cautions and limitations surround its use. The test has often been described as “assessing” brain damage, yet it is perhaps more accurate to say that it is a “screening” device for gross neurological damage. It does not provide in-depth information about the specific details and varieties of such damage. In fact, the Bender is limited to relatively severe forms of brain damage, especially in the parietal region of the right hemisphere (Black & Bernard, 1984). Thus, a patient may have significant lesions or subtle deficits that could easily go undetected if the Bender were the sole method used to assess the presence of cerebral impairment. It is more correct to say, then, that the Bender is a screening device for generalized impairment and/or right parietal problems.

Research on the original version of the Bender found a certain degree of overlap between emotional and organic indicators. For example, one of the better indicators for organic impairment is the presence of difficulties with overlapping, which has been found in the Bender records of 45% of patients with organic impairment. However, Lacks (1984, 1999) also found that overlapping difficulties occurred in the records of 26% of persons with personality disorders and 26% of those with psychosis. The degree of overlap occurring in the scores of different populations has led some reviewers (Dana, Field, & Bolton, 1983; Sattler, 1985) to seriously question the clinical usefulness of the Bender. The Bender–2 in part addresses this issue with its use of a global scoring system based on ratings of the overall quality of each of the designs rather than noting various types of specific errors. However, these global ratings are also based on various errors that might be present in the designs. For example, a rater who notices rotations or difficulty making dots will rate the quality of the design lower. It is thus likely that emotional difficulties will still have some degree of overlap with the scoring system for the Bender–2. Future research will no doubt explore this issue in further detail.

A further difficulty with the original Bender was the absence of a commonly accepted and verified coding and interpretation system. The result was that different research studies often used different systems, which made it somewhat difficult to compare their conclusions. Clinicians generally begin by learning a system of coding and interpretation but end up with their own unique, subjective approach based on clinical impressions. Although this approach may be highly workable and flexible, disagreements between “experts” can occur because of the different ways in which they approach the designs. Another difficulty related to depending on clinical impressions is the continued, unwarranted reliance on unsubstantiated and possibly incorrect clinical “lore.” Lacks (1984, 1999) presented evidence that clinicians could increase their diagnostic accuracy for organic impairment on the average of 10% to 15% by using a brief, easily learned, objective scoring system. The Bender–2 and its use of a global coding system achieves exactly this goal.

Administration

Administration of the Bender–2 includes four phases. Clients are first requested to draw the designs presented to them one at a time (the “copy” phase). The next, or “recall,” phase begins when the examiner removes the designs and the paper with the

reproductions on them, gives clients a new blank sheet of paper, and asks them to copy as many of the designs as they can from memory. The two procedures that follow are the perception and motor subtests. During the copy phase, the examiner presents the cards directly in front of the client one at a time. The series of cards presented should be appropriate for the client's age (cards 1 to 13 for children 4 to 7 and cards 5 to 16 for persons 8 and older). Clients are asked to copy each design with a pencil on a single, blank, 8.5-by-11-inch sheet of white paper that has been presented to them in a vertical position. A sharpened backup pencil should be available in case a client breaks the tip of the pencil. Pencils should include erasers. These verbal directions are taken from Brannigan and Decker (2003) and are recommended as a standard procedure:

I have a number of cards here. Each card has a different drawing on it. I will show you the cards one at a time. Use this pencil (give pencil to examinee) to copy the drawing from each card onto this sheet of paper (point to drawing paper). Try to make your drawings look just like the drawings on these cards. There are no time limits, so take as much time as you need. Do you have any questions? Here is the first card. (p. 17)

After the person has completed the first design, the next one should be presented until the entire set of designs has been reproduced. Timing should begin immediately following the presentation of the first design. This sequence should be continued and the total time noted after the last drawing has been completed. No comments or additional instructions are to be given while clients are completing the drawings. If clients ask specific questions, they should be given a noncommittal answer, such as "Do the best you can" or "Begin wherever you like."

If clients begin to count the dots on Design 6, the examiner may say, "You don't have to count the dots, just make it look like the picture." If they persist, this may show perfectionistic or compulsive tendencies, and the behavioral observation should be considered when evaluating the test results and formulating diagnostic impressions. Although examinees are allowed to pick up the cards, they are not allowed to turn them unless they are in the process of completing their drawing. If it looks as if they have turned the design and are beginning to copy it in the new position, the examiner should straighten the card and state that it should be copied from this angle. As many sheets of paper may be used as desired, although clients are presented with only one sheet initially. There is no time limit, but it is important to note the length of time required to complete the test, as this information may be diagnostically significant. An observation form is included that enables examiners to note physical and test-taking observations for each of the designs, as well as to record the scores.

After the copy phase, a memory or "recall" phase is presented. Immediately after having copied the final designs, clients are given a new sheet of paper and given the following instructions (from Brannigan & Decker, 2003):

Now I want you to draw as many of the designs that I just showed you as you can remember. Draw them on this new sheet of paper. Try to make your drawings just like the ones on the cards that you saw earlier. There are no limits, so take as much time as you need. Do you have any questions? Begin. (p. 18)

Begin timing as soon as clients begin their first drawing and stop when they have indicated that they cannot recall any more of the designs (after approximately

2 minutes). Be sure to label the sheet as “Recall Sheet” and indicate where the top of the page is with an arrow. The recall phase provides an assessment of clients’ level of short-term, visuomotor recall. Typically, adults with brain injury are not able to recall the designs as well as persons who do not have a brain injury (Lyle & Gottesman, 1977).

The Motor Test consists of a sheet of paper with a sample item and four test items. Examinees are first shown the sample item depicting a rectangle with two medium-size dots on either end. A series of smaller dots connect the two medium-size dots. Examinees are given these instructions, which are read from the top of the page of designs:

For each item, start with the largest figure. For each figure, draw a line connecting the dots without touching the borders. Do not lift the pencil, erase, or tilt the paper while drawing.

Sometimes it might be necessary to demonstrate the procedure by drawing a line between the two medium-size dots in the sample. Begin timing when the examinee begins the first design. Discontinue the procedure and note the time after the examinee either finishes the procedure or after a maximum of 4 minutes has elapsed.

This Motor Test will help detect the presence of motor problems. Another possibility is the presence of perceptual problems. Often these can be detected with the Perception Test, which depicts 10 designs, each of which is followed by 4 designs. One of these 4 is identical to the original design; the others are merely similar. According to the Bender–2 manual (Brannigan & Decker, 2003), the examiner is to say:

Look at this picture (point to the design in the first box). There is another picture that looks just like it in this row (run your finger across the first row). Circle or point to the picture that looks just like this one (point again to the designs in the box).

If needed, provide assistance for the first item. Point to each item in the row and say: Which one of these pictures looks like this one? (point again to the design in the box). (p. 19)

This procedure should be timed but should be discontinued if it takes more than 4 minutes. If an examinee takes more than 30 seconds to give a response to any of the items, the item should be discontinued and the examinee should be instructed to go to the next item.

Scoring

Scoring for both the copy and recall phases requires examiners to rate each of the designs drawn on a scale between 0 and 4 where:

- 0 = No resemblance, random drawing, scribbling, lack of design
- 1 = Slight-vague resemblance
- 2 = Some-moderate resemblance
- 3 = Strong-close resemblance, accurate reproduction
- 4 = nearly perfect (Brannigan & Decker, 2003, p. 20)

Examples are provided in the manual to assist with scoring. The scores are totaled with higher scores indicating better performance. Scores for examinees below age 8 can

range between 0 to a maximum of 52. In contrast, scores for examinees above the age of 8 can range from 0 to a maximum of 48. The raw scores are then converted to standard scores ($M = 100$, $SD = 15$) using age-referenced tables in the back of the manual.

The Motor Test is scored a 1 if the line touches both of the medium-size dots at either end of the design and does not cross any of the borders. The line can touch the border of the design but cannot go over it. In contrast, a 0 is scored if the line extends outside the box or if it does not touch both of the medium-size dots (which represent the end points). A total of 12 points is possible.

The Perception Test is scored a 1 for a correct response and a 0 for an incorrect response. Thus a total of 10 points is possible. A table at the back of the Observation Form converts both the Motor Test and Perception Test to percentile rankings.

Although there is no specific place for this on the record form, it is strongly suggested that examiners transform the completion times for the copy and recall phases into standard (z) scores and percentile ranks. Using Table D-1 on page 116 of the test manual for both the copy and recall times, examiners can transform the completion times into z scores by subtracting the mean time for the appropriate age group from the examinee's completion time, then dividing by the standard deviation listed in the manual. Note that the times will need to be converted into seconds, rather than calculated as minutes and seconds. Z scores, with a mean of 0 and a standard deviation of 1, relate directly to percentile ranks, which can be meaningful alone or in conjunction with other measures, such as the Processing Speed Index on the Wechsler intelligence scales.

Interpretation Guidelines

In order to perform well on the Bender, clients must have adequate fine motor coordination and the ability to make accurate perceptual discriminations. They must then integrate this into the actual reproduction of the design. In contrast, errors may reflect poor fine motor coordination, difficulties actually perceiving the design, problems executing the drawing itself, or difficulties integrating the perception and motor requirements (Sattler & Hoge, 2006). Difficulties with poor performance may be the result of delays in visuomotor abilities, brain dysfunction, emotional disturbance, or a combination of all of these factors. Bender performance seems to be only minimally influenced by cultural factors or processing speed (Decker et al., 2006).

The Global Scoring on the Bender-2 allows for raw scores to be converted into standard scores ranging between 40 and 160. Similar to the Wechsler intelligence scales, there is a mean of 100 and a standard deviation of 15. Thus a score of 85 on the Bender-2 means the client scored 1 standard deviation below the norm or at the 16th percentile when compared with his or her age-related peers. Brannigan and Decker (2003) stated that scores in the lower 25th percentile should signal the need for further evaluation. Often a score in the bottom 2% of the population (2nd percentile, standard score of 70) is considered to be in the "impaired" range. However, these considerations depend largely on the person's history, demographics, and level of functioning. For example, a student who has been functioning near the top of the class but who then begins to have academic difficulties and has Bender-2 scores at the 20th percentile may indeed suggest a deteriorating condition. In contrast, another student with a marginal academic record, low-average intelligence, and who similarly has a Bender-2 score in

the 20th percentile may be merely reflecting overall low-average abilities. Brannigan and Decker provided these classifications for Bender–2 standard scores:

Extremely high or extremely advanced	145–160
Very high or advanced	130–144
High or advanced	120–129
High average	110–119
Average	90–109
Low average	80–89
Low or borderline delayed	70–79
Very low or mildly delayed	55–69
Extremely low or moderately delayed	40–54

In a general way, low scores on the Bender–2 typically represent the person’s visuo-motor abilities. The listed normative and extra test comparisons represent the most obvious and clear way of making very general sense of the person’s scores. Beyond this, three major areas need to be considered in expanding on the meaning of the person’s score: (1) distinguishing between perceptual versus motor difficulties; (2) considering the meaning of design construction versus visual memory; and (3) differentiating among developmental delays, brain dysfunction, and emotional disturbance.

Perceptual versus Motor Difficulties

A useful informal guide in distinguishing between whether poor Bender–2 performance is due to perceptual versus motor difficulties is to carefully consider relevant behavioral observations, along with qualitative features of the drawings. These observations might include areas such as the client’s level of confidence, awareness of errors, completion time, and any comments that are made. The clinician might then look at specific features of the drawings, including figure size, placement, line quality, order of the designs, location of the designs, distortions, erasures, reworking, omissions, and any other unusual treatment. A number of these types of observation can be made on the Bender–2 protocol in the sections for “Physical Observations” and “Test-taking Observations.”

These observations can be useful in determining whether a client’s poor Bender–2 reproductions are the result of inadequate perception (difficulty in receiving visual informal), inadequate motor abilities (difficulty physically rendering the designs), or the integration of the two (difficulty in reproducing that which might have been accurately perceived). This distinction can sometimes be made by asking the person to evaluate the accuracy of the drawing he or she has made. If clients feel that poorly reproduced drawings are accurate, then they most likely have receptive difficulties and possibly difficulties with expression. If they recognize that their drawings were done poorly, this suggests their problem might be primarily motor/expressive. Although they might be aware of the inaccuracy of their drawings, individuals with these types of motor problems would be expected to have difficulty in correcting the inaccuracies.

The Bender–2 has formalized this differentiation through the use of the Motor and Perception subtests. If clients have a poor score on the Bender–2 but the Motor Test

was average or high, it suggests that their perceptual abilities were likely the major problem. This finding would then be confirmed by whether they had a low score on the Perception subtest. Conversely, poor performance on the Motor subtest but intact performance on the Perception subtest would suggest that the problem was due to motor abilities. When the Bender–2 copy score is low but both the Motor and Perception subtests are average or better, it is more likely that the integration of perception and motor output is the problem.

Considering Copy versus Recall Performance

A further refinement is to note the relative scores on the Bender–2 copy versus the recall phases. The previous nine-design Bender used general or “clinical” norms to indicate low, average, or high performance. Specifically, it was expected that a healthy person of average intelligence would accurately recall and construct four to five designs. An advantage of the Bender–2 is that the more global scoring allows for a much more precise calculation in which a raw score can be converted to a standard, age-related score. Additionally, different processes are assessed by the copy and recall phases. Additional information and validation for each can be obtained by asking clients for examples in their lives that reflect any difficulties (or strengths) with visual ability and memory. This may include their ability to recall who was at meetings, where they had left things in their homes, or ability to find somewhere that they had been before. If the Wechsler Memory Scale—IV was given, the clinician might make sure the Bender–2 recall score is consistent with Wechsler measures of visual memory.

A normal copy phase with a contrasting poor recall performance indicates good constructional abilities but possible problems with visual memory. This finding may be consistent with a condition such as Alzheimer’s disease or the difficulties with memory consolidation that typically follow traumatic brain injury (Brannigan & Decker, 2003).

Possible Causal Patterns

A third and final aspect of Bender–2 interpretation is inferring causal patterns related to developmental delays, brain dysfunction, or emotional disturbance. Sometimes a combination of all of these problems results in low Bender–2 performance. An important issue is that different clients might have the same score but for different reasons. For example, a poorly drawn design may result from a neurologically based processing deficit in one person; for another it may result from an emotionally based sense of disorientation. Another example may occur when a person with a perfectionistic, obsessive style but with no indication of brain damage takes considerable time to complete the drawings. Another person with documented brain damage who also takes longer than average but insists on counting each dot may be attempting to compensate for his or her impairment by developing purposeful, obsessive behaviors. Other factors that might influence Bender–2 performance are situations that might encourage faking, chronic schizophrenia, older age, or a history of substance abuse. Sattler and Hoge (2006, pp. 218–219) listed the possible reasons for poor Bender performance as including:

- visual problems;
- physiological limitations associated with illness, injury, fatigue, or muscular weakness;

- physically disabling conditions, such as low birthweight, cerebral palsy, or sickle cell anemia;
- environmental stresses;
- impulsiveness;
- inadequate motivation;
- emotional problems;
- mental retardation
- social or cultural deprivation; and/or
- limited experience.

Sometimes the presence and severity of different types of errors, along with relevant behavioral observations, can be used to form *tentative* hypotheses concerning client functioning. In particular, there are often qualitative differences in the performance of persons with lesions in different areas of the brain. Whereas patients with right-hemisphere lesions are more likely to make errors related to visuospatial abilities (e.g., rotations, asymmetry, fragmentation, unrecognizable drawings, unjoined lines), persons with left-hemisphere lesions often make drawings that are shaky (line tremors) and smaller in size, with rounded corners and missing parts (oversimplification; Filskov, 1978). However, the Bender–2 is still more generally sensitive to being lowered due to right-hemisphere difficulties. In contrast, it is likely to miss patients who have left-hemisphere lesions.

Another pattern is when clients have primary difficulties with incorrect rotations in their reproductions of the designs. This might reflect mirror reversals involved with other tasks, such as reading. In contrast, other clients might have difficulties in sequencing, which could be suggested by a poorly arranged sequence in the reproduction of their Bender–2 designs.

A useful interpretative strategy is to note and compare scores on other relevant tests. If people do poorly on the Bender–2, it would be expected that they would similarly do poorly on the Weschsler Block Design subtest. An advantage of the subtest is that careful behavioral observation can help the practitioner more fully understand clients' deficits. Clients with perceptual difficulties do poorly, primarily because they distort and misperceive the design. These difficulties are more consistent with right parietal lesions. In contrast, patients with left parietal lesions are able to correctly perceive the overall gestalt of the design, but their problem-solving style may be confused and simplistic. Other clients might be able to understand the task and perceive it correctly but still experience difficulty in actually completing the task. This dissociation between intent and actually being able to make the blocks do what they want is formally referred to as *constructional dyspraxia*. Sometimes clients with a concrete orientation to problem solving do quite poorly on Block Design, because it requires a certain degree of abstraction.

In addition, it would be expected that clients who perform poorly on the Bender–2 would also perform poorly on other drawing tests, such as drawings of a clock, human figure, or bicycle. A test like the Bender–2 is quite structured (as would be asking clients to draw a picture of a clock). In contrast, having them perform a free drawing test, such as drawing a picture of a person, is far less structured. Clients must initiate, organize,

and monitor their activity to a greater extent. As such, free drawing procedures add a different dimension to the more structured Bender–2 and Wechsler subtest tasks. Formal scoring criteria and norms can be found for clock, bicycle, and house drawings in Lezak et al. (2012) and for clock drawings in Strauss et al. (2006).

Sometimes clients have learned to compensate for visuomotor difficulties caused by CNS complications. As a result, their Bender–2 reproductions might be relatively accurate. This compensation is particularly likely if an injury is not too extensive, there was above-average premorbid intelligence, the location of the lesion is not too critical, and the injury is not recent. Clinicians sometimes can detect the possible presence of brain damage by becoming sensitized to a wide range of possible compensatory mechanisms. Koppitz (1975) listed some of these:

- Excessive length of time for completion
- “Anchoring” designs by placing a finger on them as they attempt to reproduce them
- Reproducing a design from memory after first glancing at it
- Checking and rechecking the number of dots yet still being uncertain regarding the correct number that should be included
- Rotating either the sheet of paper or the Bender–2 card itself as an aid in reproducing the design
- Designs that are quickly and impulsively drawn and then corrected with extreme difficulty
- Expressions of dissatisfaction with the poorly reproduced designs followed by repeated efforts to correct them

When screening for neuropsychological impairment using the Bender–2, it is important to be aware that many of the indicators for CNS problems are also indicators for emotional disturbance. This fact raises the serious possibility of misclassification. Thus, the results of the Bender–2 alone are rarely sufficient to make a differential diagnosis between neuropsychological impairment and emotional disturbance; additional information is needed to determine both the nature and the cause of the individual’s problems.

REPEATABLE BATTERY FOR THE ASSESSMENT OF NEUROPSYCHOLOGICAL STATUS UPDATE

History and Development

The Repeatable Battery for the Assessment of Neuropsychological Status Update (RBANS; Randolph, 2012) is a relatively brief series of 12 subtests that provide a wide selection of tasks that have been found to be sensitive to neuropsychological impairment. The Update version of the test is nearly identical to the original version of the test (Randolph, 1998), except for downward extension of the norms to adolescents as young as 12 years old. The 12 subtests are organized into five indexes, each of

which provides its own score along with a Total Score (see Table 12.4). The test takes 20 to 30 minutes to administer and can be used on persons between the ages of 12 and 89. According to Randolph (2012), the RBANS was developed to serve as a stand-alone battery for identifying and describing dementia among elderly patients. However, research has found it can be used effectively with a much wider range of patients. The RBANS was also developed to serve as a screening battery when longer, more detailed tests are neither practical nor available. As the name suggests, it can also be used to provide repeat evaluations. To reduce the impact of practice effects, four parallel forms can be administered.

Whereas the Bender–2 focuses on visuomotor abilities, the subtests of the RBANS use a wide range of procedures related to such areas as verbal skills, attention, visual memory, and visuoconstruction. Practitioners can get general information on these types of domains by reading the previous sections on the neuropsychological domains of attention, language, memory, and visuospatial functioning. The RBANS subtests themselves are variations of commonly used procedures in neuropsychology. For example, learning a list of words, repeating details from a story, and reproducing complex designs are commonly used and validated procedures. The use of variations of familiar tests should make it easy for practicing professional psychologists to adapt to and interpret many of the RBANS subtests. The procedures are also sufficiently easy for psychologists in training to learn and use. The procedures for the RBANS were, in part, selected to be sensitive to a wide range of conditions. Since the RBANS assesses different domains, it was hoped that it would be effective at noting different patterns of impairment. For example, one set of patients might present primarily with memory problems, whereas another group might primarily have difficulties with attention and verbal fluency. It was hoped that the index score patterns could be used to distinguish among various groups of patients.

In addition to assessing a wide number of discrete domains, the RBANS was designed to be brief (under 30 minutes), portable, have alternate forms, and have moderate difficulty. The moderate difficulty means that it would be sensitive to early, subtle symptoms of conditions such as dementia. A number of other procedures, such as the Mini Mental State Examination or the Dementia Rating Scale, have not been found to be particularly sensitive to the early phases or mild presentation of cognitive difficulties. Many of the other longer, more complex neuropsychological tests are sensitive to subtle cognitive difficulties but are overly challenging for many patients with neuropsychological problems.

The RBANS is an ideal screening procedure. Many acute care settings have patients who present with altered cognitive status (e.g., from anoxia, traumatic brain injury, stroke) and need a fairly brief assessment of their functioning. The patients' level of cognitive impairment needs to be evaluated quickly so that decisions for level of care or referral to other health professionals can be made. The 20- to 30-minute RBANS measures a wide range of functions and is easy to learn, and it thus represents an excellent option. In addition, patients often need to have their recovery tracked during rehabilitation. Patients with progressive diseases also need to be monitored to assist with ongoing treatment and decision making. The RBANS is also an ideal instrument for nonneuropsychologists as it is easy to learn, administer, and interpret. A Spanish-language version is also available, with normative data extrapolated by

Table 12.4 Description of the Repeatable Battery for the Assessment of Neuropsychological Status Update

Index	Subtest	Description
Immediate Memory	List Learning	List of 10 words read orally; client is requested to learn the words over four trials.
	Story Memory	Short story is read orally, client is requested to recall details over two trials.
Visuospatial/Constructional	Figure Copy	Complex picture is presented, and client is requested to reproduce it.
	Line Orientation	Client is shown one diagram with two lines radiating from it and must match the two lines with numbered lines radiating from an anchor diagram.
Language	Picture Naming	Client is requested to name pictures of various objects.
	Semantic Fluency	Client is requested to name as many words as possible within various categories (e.g., fruits, animals) in 1 minute.
Attention	Digit Span	Lists of numbers are read to the client, who is requested to recall the numbers in order.
	Coding	Client is shown rows of boxes with geometric designs and must fill in numbers below the designs that correspond to anchor boxes with both designs and numbers.
Delayed Memory	List recall	Client is requested to recall as many of the words from List Learning as possible 20 minutes after having been administered List Learning.
	List recognition	Twenty words are read, 10 were on the word list from List Learning and 10 were not; client is requested to determine if the words were on the original list or not.
	Story Memory	Client is requested to recall details from Story Memory following a 20-minute delay.
	Figure Recall	Client is requested to draw the design from Figure Copy from memory following a 20-minute delay.

Source: Adapted from C. Randolph (2012) and from Strauss, Sherman, and Spreen (2006).

empirical methods from the English-speaking normative sample. Informal versions have been developed for Japan, Hungary, France, Italy, Norway, and Russia.

Norms reported in the RBANS manual (Randolph, 2012) are comprised of 540 persons between the ages of 20 and 89 who were representative of the U.S. population based on the 1995 census data, followed by an additional group of adolescents aged 12 to 19 to extend the test downward in its 2012 update. These norms are generally used when calculating the subtest and index scores. However, education can affect cognitive performance on a wide variety of tests. As a result, Duff et al. (2003) developed a larger normative sample of 718 community-dwelling adults for the RBANS. Tables are provided that allow for conversions of subtest, index, and total scores that include both age and education (reproduced in Strauss et al., 2006, pp. 240–245). Additional norms are also available for 631 older adults (278 men and 353 females) that allow for corrections based on sex and education (Beatty, Mold, & Gontkovsky, 2003). Finally, norms for older community-dwelling African Americans are also available (Patton et al., 2003).

Reliability and Validity

Split-half reliability for the Total Score was an excellent .93, and the individual index scores ranged between .75 for the Visuospatial/Constructional Index to .88 for the Immediate Memory Index (Randolph, 2012). The standard error of measurement for the Total Score was a fairly narrow 4.06 with the bands of error for the indexes ranging between 5.36 for the Immediate Memory Index and 7.63 for the Visuospatial/Constructional Index. Specific standard errors of measurement for various age ranges in relation to the Total and Index scores are provided in the record forms and manual (Randolph, 2012). Test-retest reliability (mean interval = 38.7 weeks, $SD = 2.8$) for adults was an average of .88 for the Total Score and ranged between an average of .80 for the Attention Index and .46 for the Language Index (Randolph, 2012). For adolescents, test-retest reliability (interval range between 14 and 31 days) was .85 for the Total Scale and ranged from .63 for the Visuospatial/Constructional Index to .80 for the Delayed Memory Index (Randolph, 2012). Duff et al. (2005) similarly found test-retest reliabilities for a longer 1-year duration ranging between .51 and .83 for adults. Practice effects were noted to have been largely absent.

As mentioned, the content of the RBANS subtests is quite similar to other frequently used tests in clinical neuropsychology. In other words, the types of procedures selected have proven their validity within other forms and contexts. Correlations with external measures demonstrate good to excellent convergent and divergent validity. For example, the correlation between the WAIS-R Full Scale IQ and the RBANS Total Score was a quite high .78 (Randolph, 2012). As expected, a quite high correlation was found between the RBANS List Learning subtest and list learning on the California Verbal Learning Test—II (.70) and the RBANS Coding subtest and Coding on the WAIS-III (.83; McKay, Wertheimer, Fichtenberg, & Casey, 2007). A final example is that the RBANS Visuospatial/Constructional Index had a quite high correlation with the Rey-Osterrieth Complex Figure Test (.79; Randolph, 2012). This is not surprising since most of these RBANS indexes have subtests that are derivatives from earlier “classic” tests in neuropsychology. This similarity of test format combined with the moderate to high correlation with external measures provides good

empirical support that the RBANS indexes/subtests rely on and correlate with other well-validated procedures.

A crucial area of validity is the extent the indexes actually measure discrete functions. To a certain extent, this has been supported. For example, high correlations have been found between the Immediate Memory Index and the Delayed Memory Index (.63), both of which are measures of memory (a similar pattern to the research on the Wechsler Memory Scale). In contrast, expected low correlations were found between the other, more dissimilar, index scores ($r = .21 - .47$; Randolph, 2012). There is also evidence that some categories of patients perform in a predicted pattern, given what is known about the disorders. For example, patients with Alzheimer's disease who are known to have particular difficulties with delayed memory scored much lower on the Delayed Memory Index than patients with vascular dementia (Randolph, 2012). A further example is that patients with left-hemisphere strokes scored particularly low on the RBANS Language Index and relatively better on the Visuospatial/Constructional Index (M. C. Wilde, 2006). The opposite was true for patients with right-hemisphere strokes. However, factor analytic studies have supported a two-factor structure for the RBANS rather than the five factors indicated by the index categories (Carlozzi, Horner, Yang, & Tilley, 2008; Duff et al., 2006; King, Bailie, Kinney, & Nitch, 2012; M. C. Wilde, 2006). These two factors broadly seem to relate to memory and visuospatial/constructional ability.

Whereas the RBANS was originally intended to be sensitive to early and middle stages of dementia, it has been found to be sensitive to a wide range of other conditions. Specifically, the RBANS has been found to be a reliable and valid measure for patients with traumatic brain injury (McKay et al., 2007; Pachet, 2007); stroke (Larson, Kirschner, Bode, Heineman, & Goodman, 2005); schizophrenia (Gogos, Joshua, & Rossell, 2010; Laurent et al., 2007; Randolph, 2012; Wilke et al., 2004); heavy alcohol use (A. Green et al., 2010); multiple sclerosis (Davis, Williams, Gupta, Finch, & Randolph, 2015); and Huntington's disease, Parkinson's disease, and HIV dementia (Randolph, 2012). RBANS assessment has also been found to predict everyday functioning among patients with dementia (Freilich & Hyer, 2007), stroke (Larson et al., 2005), concussion (Moser & Schatz, 2002; Moser, Schatz, & Jordan, 2005), Parkinson's disease (Beatty, Ryder, Gontkovsky, Scott, McSwan, & Bharucha, 2003), schizophrenia (Gold et al., 2002), and traumatic brain injury (McKay et al., 2007).

Assets and Limitations

The RBANS successfully fills the niche of being a broad-band, brief, moderately difficult test that is easy to administer, score, and interpret. It has good reliability and good correlations with relevant outside measures and is effective at differentiating a variety of clinical populations, including patients with cortical and subcortical dementia, as well as right- and left-hemisphere stroke. In addition, it can accurately predict relevant aspects of real-world behavior. The subtests themselves are familiar to practitioners in the field, and the tasks seem highly sensitive to cognitive impairment. It is thus an ideal screening instrument that can be used to track both a client's improvement and deterioration. As a result of these assets, the RBANS has rapidly gained acceptance in clinical and research contexts.

Despite these assets, there are also a number of problems related to its use. The five index scores provide the appearance that a full-range neuropsychological assessment might have been accomplished. Although there is some support that the indexes do indeed measure discrete functions, this has not been supported by factor analysis. In addition, a “pattern analysis” based on the index scores needs to occur only when the differences between the index scores are quite large (typically 10–20 points). Thus minor variations should not be interpreted. Tables provided in the manual (Randolph, 2012) allow clinicians to calculate both the statistical significance of differences (Table A.1, pp. 104–105) and the frequency that the differences occur (Table A.2, p. 106). For example, it might seem that a difference of 20 points between the Visuospatial/Constructional Index and the Language Index is quite large. However, this occurs in a full 27% of the standardization sample. Similarly, scores indicating actual change in a client’s level of functioning require quite large changes in retest performance (generally at least 15 points in either direction; see Duff et al., 2005; Wilke et al., 2002). In many instances, then, an increase or decrease in scores on retesting merely represents error in measurement rather than actual client change. Strauss et al. (2006) recommended that the relatively stable Total Score and possibly Attention Index should be the preferred measures for tracking client change.

A further area of caution is to take into account factors, particularly educational, that moderate the meaning of scores. The RBANS manual (Randolph, 2012) does provide some rough corrections for education (Tables 5.4–5.8), but more precise norms have been provided by Duff et al., (2003). For example, persons with a high school education between the ages of 20 and 49 have been found to have a mean RBANS Total Score of 87.5. Thus, scores from persons with lower education should not be incorrectly inferred to indicate impairment unless the clinician has first taken the effect of education into account. Score reductions for African Americans have also been noted (Patton et al., 2003). Accordingly, clinicians should take this into account when inferring such things as cognitive impairment.

Administration

Administration instructions are clearly indicated on the Record Form and in the RBANS manual (Randolph, 2012). The stimulus material is similarly clear and easy to use. Scoring criteria are straightforward and allow examiners to summarize the results on a graph and table that lists the total score, index scores, confidence intervals, and percentiles. However, if a subtest is spoiled or cannot be given due to a disability (aphasia, motor impairment), the value of the test is limited since the related index score and Total Score cannot be calculated. In addition, modifications to accommodate a patient’s disabilities or language modifications will result in a nonstandard administration. Accordingly, the data may not represent accurate information.

Interpretation Guidelines

Similar to the Wechsler intelligence scale and Bender–2, the RBANS translates raw scores into standard scores with a mean of 100 and a standard deviation of 15. The Record Form, in combination with the manual, allows these to be described as

percentiles and descriptive classifications. Thus a standard Total Score or index score of 70 means that the examinee would have scored in the borderline range or second percentile when compared with his or her age-related peers. The meaning of a score in the borderline range may need to be modified given the client's educational level and, if sufficient information is present, ethnicity.

A useful strategy is to begin with the most general level of analysis and then proceed to more specific, qualitative information. A three-step interpretive process is recommended.

Level 1: Total Score

The Total Score is the most stable, well-validated, and general measure on the RBANS. An important consideration is the degree that it agrees with other information about the patient. Thus, a high-functioning professional who scores in the average or low-average range suggests there might have been some sort of acquired difficulty. The clinician would then need to consider if there were something in the person's history that could help to explain this score (such as a set of presenting symptoms, head injury, anoxia, or stroke). It should be noted that, since the RBANS is comprised of fluid measures that are highly sensitive to cognitive impairment, RBANS total scores may be significantly lower than something like the Wechsler intelligence scale Full Scale IQ. The Total Score can be used to monitor deterioration or improvement in a client's functioning, assuming the retesting indicates a significantly large difference (generally 15 points or more).

Level 2: Analysis of Index Scores

The five index scores represent five commonly assessed domains of neuropsychological functioning. By comparing and contrasting these domains, a patient's relative strengths and weaknesses can be evaluated. For example, one patient may have poor memory (low scores on the Immediate Memory and Delayed Memory indexes) and have quite good nonverbal abilities (high/intact scores on Visuospatial/Constructional Index). For another patient, this pattern may be reversed. These patterns may help to diagnose different conditions. In addition, inferences can be made related to the patient's everyday functioning.

Randolph (2012) provided examples and evidence that patients with Alzheimer's disease, even in the early stages, are likely to have their lowest scores in Delayed Memory (poor ability to store new information), Immediate Memory (difficulty learning new material), and Language (poor verbal fluency, difficulty with word finding). He thus concluded that the RBANS is not only sensitive to Alzheimer's disease, but that the cognitive "signature" of the disease can be detected by analyzing the index scores.

Alzheimer's disease typically affects the cortical regions and, as such, is referred to as a "cortical" dementia. Another group of dementias affect primarily the subcortical regions (Parkinson's disease, Huntington's disease, ischemic vascular disease) and are thus referred to as "subcortical" dementias. These dementias are characterized primarily by impairments of attention and visuoconstruction. This subcortical pattern would be reflected in weaker Attention and Visuospatial/Construction index performance. A formal "Cortical-Subcortical deviation score" can be calculated by subtracting the

mean for Delayed Memory and Language indexes from the mean of the Attention and Visuospatial/Construction indexes. Patients with scores above 0 can be classified with dementias that are “cortical”; those with scores below 0 can be considered “subcortical.” This classification system was found to correctly classify 37 out of 40 patients (Randolph, 2012). Fink, McCrea, and Randolph (1998) were similarly able to correctly classify 93% of Alzheimer’s disease patients and 75% of those with vascular dementia. These differences are likely to be prevalent in the early to moderate stages of the diseases. As the disorder progresses, the high and low scores would be likely to flatten out.

Another pattern of index scores may reflect either left-hemisphere (verbal abilities) or right-hemisphere (nonverbal abilities) involvement. Which hemisphere is affected might be determined by noting whether the RBANS verbal (Language) or nonverbal (Visuospatial/Constructional) indexes are relatively higher.

In interpreting these patterns, clinicians should be guided by, and strike a balance between, two major principles. The first is psychometric. Difference scores are most likely to be meaningful when the discrepancies have been found to be both significantly different (see Table A.1 in the RBANS manual; Randolph, 2012) and are fairly unusual occurrences (see Table A.2 in the RBANS manual). The second consideration is more clinical and qualitative. Specifically, the pattern of scores needs to be informed by other information about the client, including such areas as medical/psychiatric history, family patterns of illness, other tests scores, presenting problem, and behavioral observations (see the “Level 3: Qualitative Information” section).

Level 3: Qualitative Information

The 12 individual subtests comprising the RBANS have mean scores of 10 with standard deviations of 3. Since these subtests are relatively short, they would not be expected to be sufficiently reliable for individual interpretation. However, they do provide what could be considered to be qualitative information when combined with relevant history and behavioral observations. For example, a patient who seems to struggle with finding the correct word and also has low scores on the language subtests (comprising the Language Index) can reasonably be said to have “low verbal fluency” and “problems with word finding.” Accordingly, this information can be included with interpretations of the patient’s performance. If these difficulties can be supported by sources in the person’s life (e.g., spouse, children), these interpretations are further strengthened.

RECOMMENDED READING

- Groth-Marnat, G. (Ed.). (2000). *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration*. New York, NY: Wiley.
- Lacks, P. (1999). *Bender-Gestalt screening for brain dysfunction* (2nd ed.). New York, NY: Wiley.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press.
- Snyder, P. J., Nussbaum, P. D., & Robins, D. L. (Eds.) (2006). *Clinical neuropsychology: A pocket handbook for assessment* (2nd ed.). Washington, DC: American Psychological Association.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.
- Vanderploeg, R. D. (Ed.). (1999). *Clinician’s guide to neuropsychological assessment* (2nd ed.). New York, NY: Psychology Press.

BRIEF INSTRUMENTS FOR TREATMENT PLANNING, MONITORING, AND OUTCOME ASSESSMENT

Since the early 1990s, there has been an increasing demand for brief, symptom-focused instruments to assist in the delivery of mental health services. One of the major reasons for this is managed care's emphasis on cost containment and documenting treatment efficacy. Most managed care organizations also have a rather narrow, symptom-oriented focus on treatment that is quite consistent with the content of many brief clinical instruments (see Maruish & Nelson, 2014). These instruments can be used to plan, monitor, and evaluate the impact of interventions. Importantly, the use of outcome measures has been found to enhance treatment success by identifying clients who are at risk of not responding to treatment (Lambert & Hawkins, 2004).

Another factor underscoring the importance of brief, focused tests has been the extensive and continually expanding research on the outcomes of mental health interventions. Pretest and posttest measures have proliferated to the extent that there are now a multitude of options to choose from (see Antony & Barlow, 2011; Maruish, 2004). It is almost a given that training and research clinics monitor their work through tests that typically take less than 10 or 15 minutes to complete. Given the models and procedures present in the research arena, managed care organizations also expect clinicians to demonstrate that the interventions they are implementing are indeed effective (Callaghan, 2001). Monitoring outcome with such measures has proven effective and valuable for treatment outcome (Anker, Duncan, & Sparks, 2009; Kraus, Castonguay, Boswell, Nordberg, & Hayes, 2011; Shimokawa, Lambert, & Smart, 2010). As a result, in 2004, 37% of professional psychologists reported using some form of outcome measure (Hatfield & Ogles, 2004).

The role of brief instruments has expanded in parallel with the dramatic increase in the areas in which psychologists have become involved (Maruish, 2004; Stout & Cook, 1999). This has included diverse roles, such as prevention; treatment planning; clinical outcomes management; risk management; evaluation of psychoactive medication; uncovering malingering and detecting undiagnosed psychopathology; assessment of chronic pain; geriatric assessment; and behavioral dentistry. As the likelihood of using a full battery has decreased, the number of additional areas that are quite likely to use brief instruments has increased.

One of the major challenges confronting psychologists is demonstrating the financial efficacy of their services. There is considerable evidence that psychosocial interventions are cost-effective in psychotherapy as well as in general health care (Groth-Marnat & Edkins, 1996). For example, it has long been known that somatizing

patients are high overusers of the medical-surgical system. Significant cost savings can be realized by extracting them from the costly (and relatively ineffective for them) medical-surgical area and moving them into the mental health area, where they can receive brief, targeted psychotherapy (Cummings, 1991). Unfortunately, there has been little research into the potential cost savings for assessment services. Rational guidelines suggest that assessment is most likely to demonstrate financial efficacy in the areas of risk management, linking assessment and treatment, using computer-assisted assessment, targeting problems most likely to result in cost savings, use of time-efficient instruments, and focusing on domains of greatest relevance to treatment planning and outcome assessment (Groth-Marnat, 1999). Yates and Taub (2003) proposed a model of how this can be accomplished by combining measures of cost, procedures, and processes and comparing these with an analysis of outcomes.

SELECTING BRIEF INSTRUMENTS

Before selecting a brief screening instrument, there should be some consideration given to who will administer and interpret it. The majority of instruments for adults (and those included in this chapter) are self-report measures. Such measures have the advantages of reducing clinician time and obtaining clients' own perception of their difficulties. However, they also have the potential for bias by the client's perceptions and are potentially subject to under- or overreporting. Other instruments are administered by professional psychologists or allied health professionals, such as primary care physicians, nurses, or clinical social workers (Bufka, Crawford, & Levitt, 2002; Maruish, 2004). On other occasions, a significant other person in the patient's life, such as a parent or spouse, completes the instrument.

The ability of an instrument to assist in planning and outcome assessment is particularly relevant for selecting brief instruments. First, brief instruments should not take longer than 15 minutes to complete (and preferably less time). In addition, they should typically be directly relevant to treatment planning and outcome assessment. In contrast, a full-battery approach before psychotherapy often provides a large amount of descriptive information, but most of this information is not directly applicable to treatment planning. A further frequent essential quality is that the instruments be useful for screening purposes. For example, the Beck Depression Inventory—II (BDI-II) can be used for each of the preceding purposes. A physician might use it to detect the possible presence of depression; a psychologist might administer it to determine the baseline level of depression severity for a client; and repeated administration could then be used to determine the effectiveness of interventions targeted to treat the depression. A managed care company would be particularly interested in this process to monitor quality control over the treatments it reimburses.

In addition to time efficiency and relevance to treatment planning and outcome evaluation, brief instruments should also be relevant to various target groups (F. Newman et al., 1999). For example, specialized variations on the BDI have been developed for children and geriatric populations. The BDI-II has also been found to be valid in the assessment of African Americans (Grothe et al., 2005) and Hispanic populations (Wiebe & Penley, 2005). Brief instruments should ideally be usable and understandable

not only by the therapist but also the client, significant others in the client's life, insurance companies, and researchers. Thus, they should be clear and direct enough such that they can be understood by both a professional and a nonprofessional audience. Because such instruments are often given over several different administrations, they should be sensitive to clinically important levels of change. As with any psychological test used by clinicians, they should also have adequate psychometric properties. Finally, interpreting the results should be uncomplicated, and the construct should be clear enough to enable feedback to the client or other relevant persons.

In addition to the many time-honored instruments, such as the BDI, instruments have been designed specifically for treatment planning and patient tracking. For example, the Outcome Questionnaire (OQ-45; Lambert et al., 1996) is a 45-item, self-report instrument that requests clients to rate various areas on a 5-point Likert-type scale. It can be used as an overall measure of client functioning, to establish a baseline, to assist with treatment decisions, and to assess common symptoms (e.g., stress, *DSM-5* codes). The results are organized around level and type of symptom distress, interpersonal relations, and relative satisfaction with social role. The Butcher Treatment Planning Inventory (Butcher, 1998) is a 210-item, self-report inventory that assesses issues and challenges that might be particularly relevant to treatment. The scales are organized around validity, treatment issues, and current symptoms. For example, the validity scales measure the extent to which the client has an overly virtuous presentation of self or tends to be close-minded with regard to his or her difficulties. Treatment issues include areas such as somatization of difficulties, low expectations regarding treatment, and narcissism. A final example is the Systematic Treatment Selection (STS) model (Beutler, Clarkin, & Bongar, 2000; Groth-Marnat, Gottheil, Liu, Clinton, & Beutler, 2008), which includes a software package (Beutler & Williams, 1999) and a clinician rating form (STS Clinician Rating Form; Fisher, Beutler, & Williams, 1999). Although the preceding instruments show considerable promise, they have not been used or tested as widely as many other instruments.

Each of the three instruments selected for this chapter fulfills the criteria required for treatment planning, monitoring, and outcome assessment. All are time-efficient and directly relevant to treatment planning, can be used to evaluate outcome, are effective as screening instruments, are relevant for a wide range of target groups, and are sensitive to change, and the constructs and information they provide are sufficiently clear so that feedback is easy to give. Each also has demonstrated its effectiveness in accurately assessing various ethnic and cross-national groups. As a result of these qualities, they have become the most extensively used brief instruments in both clinical practice and research.

SYMPTOM CHECKLIST–90–R AND BRIEF SYMPTOM INVENTORY

The Symptom Checklist–90–R (SCL-90-R; Derogatis, 1994) and its shortened version, the Brief Symptom Inventory (BSI; Derogatis, 1993), are ideally suited to quickly assess a client's type and severity of self-reported symptoms. It should not be regarded as a personality measurement; it is more an assessment of the current level of a variety of symptoms as experienced over a 1-week interval. The SCL-90-R was derived from the

earlier Hopkins Symptom Checklist (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974), which in turn had its origins in the much earlier Woodworth Personal Data Sheet (Woodworth, 1918). As the name suggests, the SLC-90-R consists of a series of 90 descriptions of symptoms that a client rates in terms of their severity (ranging from 0 = Not at all to 4 = Extremely). A sixth-grade reading level is required, and it usually takes between 12 and 15 minutes to complete. The symptoms are scored around nine different dimensions (e.g., Somatization, Obsessive-Compulsive) and three global indexes (e.g., Global Severity Index). The BSI is a short form of the SLC-90-R composed of 53 of the SLC-90-R items, and it provides scores on the same symptom dimensions and global indexes.

Scores on the SLC-90-R are transferred onto a profile sheet displaying the nine symptom dimensions and three global indexes. Similar to the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and many other measures, each score has a mean of 50 and a standard deviation of 10. One of the unique features of the SLC-90-R is that these scores can be compared with and plotted based on four normative groups:

1. Norm A: *Psychiatric outpatients* ($N = 1,002$; 425 male, 577 females), approximately two-thirds of whom were White, and the entire sample was slightly skewed toward the lower end of the socioeconomic scale.
2. Norm B: *Nonpatients* ($N = 1,000$; 494 males, 480 female), representing a stratified random sample from a large U.S. eastern state.
3. Norm C: *Psychiatric inpatients* ($N = 313$; two-thirds female), of whom 55.7% were White, 43.6% Black, with a mean age of 33.1.
4. Norm E: *Nonpatient adolescents* ($N = 806$; 60% females, 40% males), aged between 13 and 18 ($M = 15.6$) from two schools and composed primarily of middle-class Whites.

If clinicians wish to make comparisons with a nonpatient group, they can use Norm B. In other situations, it might be advantageous to compare a person with either an outpatient or an inpatient reference group. These norms, combined with the wide diversity of validity studies, suggest that the SLC-90-R can be used with a wide variety of respondents, including medical patients, adolescents, community nonpatients, cross-cultural/national groups, and inpatients and outpatients. It is also available in more than 26 languages, and computer scoring, administration, and interpretation programs are available.

Scoring and normative comparisons for the BSI follow similar procedures as for the SLC-90-R. The norms used on both scales are the same for psychiatric outpatients, psychiatric inpatients, and nonpatients. The BSI, however, has a larger adolescent normative base composed of 2,408 middle-class students (58% White, 30% Black, 12% other) between the ages of 13 and 19 ($M = 15.8$) from six different schools. Additional norms have been developed and published separately for older adults (Hale, Cochran, & Hedgepeth, 1984) and adolescent students (Canetti, Shalev, & De-Nour, 1994).

Reliability and Validity

The reliability of the SLC-90-R has consistently been good. The manual reports that internal consistency for the nine symptom dimensions based on psychiatric outpatients ranged from a low of .79 for Paranoid Ideation to a high of .90 for Depression. Internal consistency for “symptomatic volunteers” was slightly lower and ranged from a low of .77 for Psychoticism to .90 for Depression (Derogatis & Savitz, 1999). Test-retest reliability over a 1-week interval ranged from a low of .78 for Hostility to a high of .90 for Phobic Anxiety. Most coefficients were in the mid-80s. As expected, test-retest reliability was slightly lower over a 10-week interval and ranged between .68 for Somatization to .83 for Paranoid Ideation (Derogatis, 1994; Derogatis & Savitz, 1999).

Reliability for the BSI is similar although, as would be expected for a short form, slightly lower than the SLC-90-R. Internal consistency ranged between .71 for Psychoticism and .85 for Depression (Derogatis, 1993). Similarly, internal consistency for bereaved parents ranged between .74 for Psychoticism to a quite high .97 on the Global Severity Index (L. C. Johnson, Murphy, & Dimond, 1996). Test-retest reliability over a 2-week interval ranged from a low of .68 for Somatization to a high of .91 for Phobic Anxiety (Derogatis, 1993). One noteworthy feature was that the test-retest reliability was a quite high .91 for the Global Severity Index, indicating that it is a stable measure over time. This is particularly important, given that the BSI (and SLC-90-R) Global Severity Indexes are frequently used over repeated administrations to monitor treatment and evaluate its outcome. Stability over a 3-year interval (measures made every 6 months) among a group of adults with schizophrenia was also found to be quite high (Long, Harring, Brekke, Test, & Greenberg, 2007).

Well over 1,000 studies have been done investigating the validity of the SLC-90-R. For example, both MMPI and General Health Questionnaire (GHQ) measures were found to converge with expected dimensions on the SLC-90-R and diverge with other expected measures (Derogatis, 1994; Schmitz, Kruse, Heckrath, Alberti, & Tress, 1999). The SLC-90-R Depression dimension has been found to have a high correlation (.80) with the Beck Depression Inventory (Peverly & Fairburn, 1990) and to detect depression equally as effectively (Choquette, 1990).

Expected SLC-90-R profiles have been found for a variety of diagnostic groups, including depression and other affective disorders, anxiety, panic, sexual dysfunction, and substance abuse (Derogatis, 1994; Derogatis & Savitz, 1999; Prinz et al., 2013). However, other studies have questioned the divergent validity of the SLC-90-R dimensions, and authors have suggested that it be used only as a general indicator of distress (Cyr, McKenna-Foley, & Peacock, 1985; Elliott et al., 2006; Vassend & Skrandal, 1999). This controversy is strongly apparent in the findings related to factor structure. On one hand, Derogatis (1994) has reported that factor-analytic research has, with the exception of the Psychoticism dimension, matched the various dimensions of the SLC-90-R. In contrast, other research has reported anywhere from one to six factors depending on the type of population that has been studied (Cyr et al., 1985; Hayes, 1997; Piersma, Boes, & Reaume, 1994; Vassend & Skrandal, 1999).

For example, Vassend and Skrondal (1999) generally found a four-factor solution, but this varied depending on the gender and level of negative affect of the sample. They concluded that there was a “profound structural indeterminacy problem” (p. 685). Cyr et al. (1985) added that the factor structure becomes particularly uncertain when evaluated beyond the boundaries of neurotic outpatients. In contrast, Hayes (1997) found support for a six-factor solution with college students (but not for the nine dimensions listed on the SLC-90-R). Urbán et al. (2014) found that although a two-factor solution better fit the data, the nine-factor solution that represents the current SLC-90-R scales was adequate as well. Interestingly, Paap et al. (2012) found that specific characteristics related to whether a unidimensional (the Global Severity Index) or multidimensional (subscales) interpretation was more valid. The major predictor was self-reported level of distress, such that those clients with low distress had little differentiation between factors/subscales, whereas those clients with more self-reported distress were more differentiated on the scales. These differences in subject population may account for a great deal of the conflicting findings about the factor structure of the SLC-90-R and BSI.

Although the factor structure of the SLC-90-R has been equivocal, research assessing the sensitivity and specificity for various disorders has been generally supportive. For example, the SLC-90-R detected relevant symptoms among individuals with bulimia with a sensitivity of 77% and specificity of 91% (Peveler & Fairburn, 1990). Similar levels of sensitivity (72%) and specificity (87%) for detecting psychological difficulties related to diabetes were also noted. High scores on the Hostility, Paranoid Ideation, Somatization, and Obsessive-Compulsive dimensions were able to detect the presence of Cluster A (paranoid, schizoid, schizotypal) and Cluster B (antisocial, borderline, histrionic, narcissistic) personality disorders with a quite high sensitivity of 89% and even higher specificity of 97% (Starcevic, Bogojevic, & Marinkovic, 2000). As would be expected given research on a one-factor solution, the SLC-90-R has been found to effectively detect the general level of distress a person is experiencing (Derogatis, 1993, 1994). The SLC-90-R has also been found to be responsive to clinically significant change (Elliott et al., 2006; Schmitz & Hartkamp, 2000) as well as levels of distress for persons with brain injury (Hoofien, Barak, Vakil, & Gilboa, 2005; Westcott & Alfano, 2005) and screening for possible comorbidity among alcohol-abusing populations (Benjamin, Mossman, Graves, & Sanders, 2006). In contrast to this favorable research on the diagnostic utility of the SLC-90-R, the Psychoticism dimension was not able to discriminate between psychotic and nonpsychotic patients (Stukenberg, Dura, & Kiecolt-Glaser, 1990). This, in combination with low internal consistency, indicates that the Psychoticism dimension seems to have the weakest psychometric properties of all the SLC-90-R scales and should thus be interpreted with great caution.

Validity for the BSI is in part supported by the high correlations between the SLC-90-R and BSI dimensions, which range from a low of .92 for Psychoticism to a high of .98 for Paranoid Ideation (Derogatis & Savitz, 1999). This is sufficiently high so that research on the SLC-90-R not only supports the BSI, but the two tests can be used as alternate forms for each other. Additional studies support the sensitivity of the BSI to distress and suggest that it can be used to track the outcomes of various interventions (Derogatis, 1993; Derogatis & Savitz, 1999). For example, screening

of recently diagnosed patients with cancer indicates that the BSI was sensitive to varying levels of distress based on ratings using outside criterion measures (Zabora, Smith-Wilson, Fetting, & Enterline, 1990). Similarly, elevated scores on the BSI have been found for bereaved parents (L. C. Johnson et al., 1996). Ratings by experienced clinicians of the level of distress experienced by clients have also been found to have moderate correlations with the expected dimensions on the BSI (Morlan & Tan, 1998). However, correlations between the BSI and client self-ratings of level of satisfaction with psychotherapy were not correlated (Pekarik & Wolff, 1996).

Use with Diverse Populations

The SLC-90-R has consistently been found to be a reliable and valid instrument in a wide number of cross-cultural contexts. For example, Martinez, Stillerman, and Waldo (2005) concluded that the SLC-90-R measures symptom-related change in a similar way for both White and Hispanic populations. It has also been used successfully to assess level of trauma-related psychopathology in Chinese (Wang et al., 2000) and Vietnamese populations (Hauff & Vaglum, 1994), as well as general distress in: Mexico (Cruz Fuentes, Bellow, Garcia, Macías, & Chavez Balderas, 2005); Spain (Caparrós-Caparrós, Villar-Hoz, Juan-Ferrer, & Viñas-Poch, 2007); Denmark (Olsen, Mortensen, & Bech, 2004); Germany (Essau, Groen, Conradt, Turbanisch, & Petermann, 2001; Schmitz et al., 2000); Hungary (Unoka et al., 2004); Greece (Donias, Karastergiou, & Manos, 1991); Japan (Tomioka, Shimura, Hidaka, & Kubo, 2008); and Korea and Korean immigrants in America and Canada (Noh & Avison, 1992). It has also performed well for Zulu speakers in South Africa (Shanahan, Anderson, & Mkhize, 2001). A study in Italy, however, suggested that the measure be used with caution, as the factor structure did not replicate well (Prunas, Sarno, Preti, Madeddu, & Perugini, 2012).

The normative groups provided in the manual include a reasonable proportion of African Americans, especially for the psychiatric inpatient (43.6%) and outpatient (32.6%) populations. As noted, there are also adolescent norms, and the SLC-90-R has been used to document the extent and longitudinal stability of symptoms among older adult populations (Agbayewa, 1990; Levenson, Aldwin, Bossé, & Spiro, 1988).

Interpretation

In many ways, “interpreting” the SLC-90-R and BSI is straightforward because the data are descriptive rather than representative. In other words, overall severity of a client’s symptoms can be assessed through the degree of elevation on the Global Severity Index. Similarly, the severity by which a client is sensitive to the criticisms of others can be gauged by the relative elevation of the Interpersonal Sensitivity dimension. However, clinicians may also wish to extend beyond these straightforward descriptions based on their clinical knowledge. For example, a person scoring high on Interpersonal Sensitivity is likely to exaggerate criticisms, ruminate over these criticisms, experience irrational thoughts, have low self-esteem, and be low in assertiveness. The severity of a client’s condition can be further investigated by taking into account additional data. There may also be patterns of elevations that are consistent with various personality

disorders. Avoidant personalities, for example, would be expected to have high scores on Interpersonal Sensitivity, Anxiety, and possibly Phobic Anxiety. In contrast, histrionic personalities would be likely to have elevations on Somatization. These conceptual links can be used as beginning points for further investigation to see if the person does or does not have the suggested personality styles. However, these “interpretations” should be considered more as hypotheses, given the questionable independence of the nine SCL-90-R/BSI dimensions.

Interpretation can begin with the global indexes and then proceed to the dimensional and symptom/item level. Accordingly, the following information is a listing of and elaboration on the meanings of the elevations in categories under these three general groupings (adapted from Derogatis & Savitz, 1999).

Global Indexes

Global Severity Index (GSI) The GSI is a combined rating that takes into account the intensity of experienced stress along with the number of reported symptoms. As such, it is the best single indicator of distress and should be used when a single measure is appropriate. A general rule of thumb is that a *T* score above 63 suggests the presence of a clinically significant level of psychological difficulties.

Positive Symptom Distress Index (PSDI) The PSDI is an average rating for all symptoms that have been endorsed. Thus, it is a measure of symptom intensity (rather than merely the number of symptoms endorsed).

Positive Symptom Total (PST) Whereas the PSDI is a measure of symptom severity, PST represents the number (or breadth) of symptoms. Thus, clients could theoretically have a low PSDI, indicating that the symptoms they have are not particularly troubling, but might have a high PST, indicating that they had a wide, potentially complex array of symptoms.

Symptom Dimensions

Somatization (SOM) An elevation on SOM indicates that distress is experienced primarily through concerns related to actual, amplified, or imagined physical dysfunction. Complaints might be focused on cardiovascular, gastrointestinal, respiratory, gross musculature, or other bodily areas (note responses to actual items). Pain and anxiety are both likely to be present as well, thereby amplifying any physiologically based disorders. Interventions might involve increasing a client’s awareness of how he or she uses somatization as a coping mechanism, combined with alternative methods of coping, such as stress management, social skills training, hypnosis, or biofeedback.

Obsessive-Compulsive (O-C) The O-C dimension focuses on impulses, thoughts, and actions that are irresistible, repetitive, unwanted, and experienced as beyond the person’s control. Some of the items also refer to more general cognitive performance deficits (e.g., the person’s mind is going blank or he or she has trouble concentrating).

Interpersonal Sensitivity (I-S) High scores on I-S indicate clients have considerable discomfort in interpersonal situations. They have negative expectations regarding relationships and are self-conscious. When they compare themselves with others, they

typically feel inferior and thus experience self-doubt and inadequacy. Crucial to any intervention is a supportive therapeutic relationship, perhaps combined with cognitive restructuring and assertiveness training.

Depression (DEP) Elevations on DEP indicate the person is experiencing a range of depressive symptoms. These might include loss of pleasure, dysphoria, loneliness, crying, withdrawal, pessimism, sleep disturbance, alteration in appetite, poor motivation, and low energy (check individual items). There may also be the presence of suicidal ideation and other cognitions consistent with depression.

Anxiety (ANX) The ANX dimension focuses on the presence of apprehension, nervousness, trembling, and dread. High levels of anxiety may or may not be consistent with panic attacks. Physiological components of anxiety, including rapid heart rate, tension, and restlessness, are also likely to be present. Possible interventions include relaxation training, meditation, stress management, assertiveness training (and other forms of skills training), and exercise.

Hostility (HOS) Persons scoring high on HOS experience resentment, irritability, aggression, and, possibly, rage. Accordingly, anger management might be an appropriate recommendation.

Phobic Anxiety (PHOB) The PHOB dimension focuses on the presence of excessive and irrational fear related to a person, place, object, or situation. The individual might report a fear of open places, anxiety when traveling away from familiar areas, or fear of having a panic attack. Although the title of the dimension appears to be related to phobias, most of the actual items reflect the more pathological aspect of phobias to the extent that high scores may reflect agoraphobia or panic attacks rather than merely specific phobias. Interventions can be focused on the areas of greatest anxiety and might include flooding, graded exposure, relaxation training, hypnosis, and cognitive restructuring.

Paranoid Ideation (PAR) Items in the PAR dimension tap into the key dimensions of paranoid thought, including hostility, projection, grandiosity, suspiciousness, and a need for control based on a fear of losing independence. Delusions may also be present and are reflected in items related to fears of being watched, talked about, or not being given credit for achievements.

Psychoticism (PSY) High scores on the PSY dimension reflect a person who is extremely withdrawn and isolated and may be experiencing core symptoms of schizophrenia, including hallucinations (e.g., hearing voices, thought broadcasting) and thought control. Scores can be seen as being on a psychoticism continuum ranging from minor levels of interpersonal alienation to a full display of severe psychotic symptoms.

T scores above 63 on two or more of the preceding dimensions suggest that the person has clinically significant levels of psychological distress. Some of the SCL-90-R items are not scored on the preceding dimensions but may be potentially important

symptoms. These include poor appetite, sleep disturbance, fear of dying, overeating, early-morning awakening, difficulties with sleep maintenance, and guilt. These can be noted to obtain additional information. Researchers have developed additional scales that may be used in the future to extend interpretation (see the SLC-90-R Mania Scale; E. E. Hunter et al., 2000; the SLC-90-R Adult ADHD Scale; Eich et al., 2012, and factor-based scales for college students; Hayes, 1997).

Symptom Level/Item

Additional information can be obtained by noting the content of the individual items the client has endorsed. For example, items on the Depression dimension can provide specific information related to the person's depression. Importantly, the presence of suicidal ideation (i.e., items related to ending the person's own life) can be noted and should then be followed up by more in-depth assessment for risk of self-harm. The presence and extent of possible vegetative symptoms (e.g., low energy, sleep problems, loss of sexual energy) can also be noted; this may have implications for various treatment recommendations. Items that the client has answered either "quite a bit" or "extremely" can be considered critical. These should be given particular attention for assessment, treatment planning, and establishing a relevant baseline and outcome to treatment.

BECK DEPRESSION INVENTORY–II

The Beck Depression Inventory–II (BDI-II) is an extremely brief, 21-item self-report measure that assesses symptoms primarily related with depression. Clients are asked to think over the last 2 weeks and determine how they have been feeling on a 4-point scale, ranging from 0 to 3 (different items have different meanings attached to the 4 points in the scale). The BDI-II generally takes 5 minutes or less to complete, and scoring is extremely quick and straightforward. It has been validated for both adults and adolescents, and it provides an excellent screening for depressive symptomatology.

The Beck Depression Inventory (BDI) was introduced in 1961 by A. T. Beck, Ward, Mendelson, Mock, and Erbaugh, was revised in 1971, and was copyrighted in 1978 (A. T. Beck, Rush, Shaw, & Emery, 1979). Although the later version, referred to as the BDI-1A, involved a clarification and modification of the items, the two versions were found to be highly correlated (.94; Lightfoot & Oliver, 1985). The BDI underwent a further and major revision in 1996 (BDI-II) to include a wider range of symptoms (A. T. Beck, Steer, & Brown., 1996). By so doing, it became more congruent with *DSM-IV* diagnostic criteria for depressive disorders. Four of the items were replaced to reflect symptoms consistent with more severe depression (agitation, worthlessness, concentration difficulty, and loss of energy). A further two items were revised to better reflect decreases in appetite and sleep. In addition, many of the other items were reworded.

Comparisons between the BDI/BDI-1A and the BDI-II indicate that clients are likely to endorse one to two more items/symptoms on the BDI-II when compared with the earlier BDI and BDI-1A (A. T. Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Steer, Rissmiller, & Beck, 2000). More symptoms are likely to be endorsed toward the higher ranges of depression (three or more items/symptoms) than the lower ranges.

Using an outpatient sample, BDI-IA/BDI-II correlations were .84, and the mean total scores were slightly higher for the BDI-II than for the BDI-IA (21.63 versus 18.15; A. T. Beck et al., 1996). Correlations between the BDI and BDI-II for a university population indicated a slightly higher correlation of .92 (Dozois et al., 1998). Despite the slightly higher scores on the BDI-II, this information indicates that the BDI-II is sufficiently comparable to its predecessors such that, with appropriate caution, much of the research on the BDI/BDI-IA can be generalized to the more recent BDI-II.

The BDI-II and its predecessors have been widely used for the assessment of depression among psychiatric patients (Camara, Nathan, & Puente, 2000; C. Piotrowski, 1996; Steer, Ball, Ranieri, & Beck, 1999; Steer et al., 2000), as well as depression in nonclinical adults (A. T. Beck et al., 1996; Steer, Beck, & Garrison, 1986). The BDI has been found to detect depression as effectively as longer and more costly structured interviews (Stukenberg, Dura, & Kiecolt-Glaser, 1990). The popularity of this instrument is amply demonstrated in that, in the nearly 50 years since its introduction, well over 1,000 research studies have been performed either on or using it.

The items in the BDI were originally derived from observing and summarizing the typical attitudes and symptoms presented by depressed psychiatric patients (A. T. Beck et al., 1961). A total of 21 items related to various symptoms were included, and, when completing the inventory, respondents are requested to rate the intensity of these symptoms on a scale from 0 to 3. Typical questions relate to areas such as sense of failure, guilt feelings, irritability, sleep disturbance, and loss of appetite. The inventory is self-administered and takes from 5 to 10 minutes to complete. A fifth- to sixth-grade reading level is required to adequately comprehend the items. The total possible range of scores extends from a low of 0 to a theoretical high of 63. However, the most severe levels of depression are reflected by scores of 40 or 50. More typically, clinically depressed or nondiagnosed but depressed nonclinical populations score in the 14 to 28 range (A. T. Beck et al., 1996).

Reliability and Validity

Although the original BDI presented with good internal consistency (ranging from .73 to .92 with a mean of .86; A. T. Beck, Steer, & Garbin, 1988) and adequate test-retest reliability (ranging from .48 to .86, depending on the interval between retesting and type of population; Beck et al., 1988), repeat administrations over 7 weeks at one administration per week using university students indicated a 40% decline in scores (Ahava, Iannone, Grebstein, & Schirling, 1998). Research with the BDI-II has consistently found high internal consistency, ranging from .89 to .94, even when using a variety of populations (Arnau, Meagher, Norris, & Bramson, 2001; A. T. Beck et al., 1996; Dozois et al., 1998; Steer et al., 1999, 2000; Subica et al., 2014). Test-retest reliability over a 1-week interval was .93 (A. T. Beck et al., 1996).

Evaluation of content, concurrent, and discriminant validity, as well as factor analysis, has generally been favorable for the BDI-II. The content of the BDI items was derived by consensus from clinicians regarding symptoms of depressed patients combined with considerations related to the various *DSM-IV* categories for the diagnosis of depression. Concurrent validity is suggested by high to moderate correlations with clinical ratings for psychiatric patients (A. T. Beck et al., 1996). In addition, moderate

correlations have been found with similar scales that also rate depression, such as the Hamilton Psychiatric Rating Scale for Depression (.71), Beck Hopelessness Scale (.68; A. T. Beck et al., 1996), and the Depression Anxiety Stress Scale (.88; Osman et al., 1997). The BDI-II has been able to discriminate psychiatric from nonpsychiatric populations (A. T. Beck et al., 1996) as well as discriminate the level of adjustment in psychiatric populations (Arnau et al., 2001; A. T. Beck et al., 1996). The ability of the BDI-II to discriminate between primarily anxiety as opposed to primarily depressive disorders is supported in that BDI-II scores were more highly correlated with the Hamilton Psychiatric Rating Scale for Depression (.71) compared with the Hamilton Rating Scale for Anxiety (.47). Similarly, Steer et al. (2000) found higher correlations between the BDI-II and the SLC-90-R Depression dimension (.89) than the SLC-90-R Anxiety dimension (.71).

A number of factor-analytic studies indicated that the BDI is composed of a Noncognitive (or somatic-affective) factor comprising contents related to somatic aspects of depression (loss of energy, changes in sleep patterns, crying) and a Cognitive factor related to self-reported thoughts clients use to describe their attitudes toward themselves and their depression (self-dislike, suicidal thoughts, thoughts of worthlessness; A. T. Beck et al., 1996). These factors have been found to be consistent among various samples, including college students (A. T. Beck et al., 1996), adolescents (Steer, Kumar, Ranieri, & Beck, 1998), older adults (Steer et al., 2000), primary care medical patients (Arnau et al., 2001; Hiroe et al., 2005), geriatric inpatients (Steer et al., 2000), and outpatients with clinical depression (Steer et al., 1999). Most comparisons between the BDI and BDI-II indicate that the factor structure on the BDI-II is more clearly defined, suggesting it is a slightly superior instrument (Dozois et al., 1998).

Despite the frequent finding of two factors on the BDI-II, other research has favored a three-factor solution (Vanheule, Desmet, Groenvynck, Rosseel, & Fontaine, 2008). For example, Osman et al. (1997) found a factor structure composed of Negative Attitude, Performance Difficulty, and Somatic Elements using a sample of undergraduates. A further three-factor structure comprised of affective, cognitive, and somatic components was found by Vanheule et al. (2008). L. C. Ward (2006) analyzed six data sets from previous studies. He found a three-factor solution comprised of a General factor combined with Cognitive and Somatic factors. Bühler, Keller, and Läge (2014) concluded that the typical factor structure was inadequate and added a factor for activation level of depressed symptoms. Vanheule et al. (2008) noted that that, across different samples, the Cognitive and Somatic factors were unstable. In other words, the items that loaded on these two factors varied according to different samples that were used. The advantage of the General factor is that it provided most of the “binding” between both factors (internal consistency) and was consistent across the various samples.

Because of the disagreements just mentioned and the general instability of specific factors across samples, more recent literature emphasizes the primacy of a unidimensional interpretation of the BDI-II, with a single general factor, as additional specific factors did not add significant information to models (Brouwer, Meijer, & Zevalkink, 2013; Quilty, Zhang, & Bagby, 2010; Reise, Moore, & Haviland, 2010; Subica et al., 2014). Separating out the cognitive and noncognitive factors may not be especially

useful, as the factors tend to be highly correlated (Ward, 2006) and, indeed, the symptoms themselves are most often overlapping in depressed individuals (Blatt, 2004).

Use with Diverse Groups

The BDI-II appears to be an appropriate measure for various ethnic and cross-cultural groups, including African Americans and Hispanics. Specifically, scores do not seem to be different among ethnic groups (A. T. Beck et al., 1996), and the factor structure has been found to be quite similar (Grothe et al., 2005; Penley, Wiebe, & Nwosu, 2003). Whisman, Judd, Whiteford, & Gelhorn (2013) found factorial invariance in a large, diverse sample of college students across race, ethnicity, and even gender. In addition, Spanish translations have been found to maintain assessment utility (Penley et al., 2003; Wiebe & Penley, 2005), as have versions for Portuguese (Campos & Gonçalves, 2011); Chinese (Chang, 2005); use in Monterrey (de la Rubia, 2013); Persian (Ghassemzadeh, Mojtabai, Karamghadiri, & Ebrahimkhani, 2005); Greek (Giannakou et al., 2013); Indonesian (Ginting, Näring, van der Veld, Srisayekti, & Becker, 2013); Japanese (Kojima et al., 2002); use in South Africa (Kagee, Nel, & Saal, 2014); and use with Arab populations (Al-Musawi, 2001). Nuevo et al. (2009) found evidence for validity across multiple European countries as well. The BDI-II has also been found to be effective in assessing the presence and extent of depression in adolescents, although a cutoff of 21 or higher has been recommended (Beltrán, Freyre, & Hernández-Guzmán, 2012; Dolle et al., 2012; Kumar, Steer, Teitelman, & Villacis, 2002; Osman, Kopper, Barrios, Gutierrez, & Bagge, 2004). At the other extreme, the BDI-II can be used effectively to assess the level and extent of depression among older adults (Segal, Coolidge, Cahill, & O'Riley, 2008; Steer, Rissmiller, & Beck, 2000).

Interpretation

These scores can be used to indicate the general level of depression:

0–13	No or minimal depression
14–19	Mild
20–28	Moderate
29–63	Severe
Below 4	Possible denial of depression, faking good; lower than usual scores even for nonclinical adults

Scores significantly above even those of severely depressed persons suggest potential exaggeration of depression, a possible characteristic of histrionic or borderline personality disorders. Significant levels of depression are still possible. Arnau et al. (2001) found that a cutoff score of 18 correctly classified 92% of patients with major depressive disorder.

An ipsative interpretation of BDI responses can be used to specify irrational beliefs and relevant symptoms that are likely to be related to a person's depression. Identification of these beliefs and symptoms can be useful in specifying areas that need to be

worked on in therapy. Any of the next 21 areas (A. T. Beck et al., 1996, p. 5) can be assumed an area of difficulty if it receives a score of 3:

1. Sadness
2. Pessimism
3. Past failure
4. Loss of pleasure
5. Guilty feelings
6. Punishment feelings
7. Self-dislike
8. Self-criticalness
9. Suicidal thoughts or wishes
10. Crying
11. Agitation
12. Loss of interest
13. Indecisiveness
14. Worthlessness
15. Loss of energy
16. Changes in sleeping pattern
17. Irritability
18. Changes in appetite
19. Concentration difficulty
20. Tiredness or fatigue
21. Loss of interest in sex

One specific area to be alerted to is the potential for suicide, which can be indicated by strong endorsements (2 or 3) on items 9 (suicidal thoughts or wishes) and 2 (pessimism). Whereas the level of depression (based on total score) and presence of specific item endorsement can assist in suggesting the presence of a formal *DSM-5* disorder, a definitive diagnosis would still need to be made based on a more thorough review by a clinician.

STATE TRAIT ANXIETY INVENTORY

A client's level of anxiety is one of the most crucial dimensions to assess, both for treatment planning and to establish the impact of interventions. The State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is ideally suited for this purpose because it is a brief (40-item), self-report inventory that is easy to understand (with a sixth-grade reading level) and is sensitive to transitory episodes of anxiety (states) as well as more stable personality features that predispose a client to experiencing more chronic levels of anxiety (traits). Respondents are asked to respond on a 4-point scale ("Almost Never" to "Almost Always" for the trait scale and "Not at

All” to “Very Much So” for the state scale) to items related to their current anxiety and their general state of worry. The STAI is currently the most frequently used measure of anxiety with over 8,000 studies available in the literature. Research has evaluated its use in relation to phobias, test anxiety, panic, generalized anxiety, and the impact of specific types of treatment, such as cognitive behavior therapy, systematic desensitization, relaxation, and rational emotive therapy (Spielberger, Sydeman, Owen, & Marsh, 1999). It has also been used extensively in cross-cultural research and has been translated into more than 60 languages and dialects.

Construction of the STAI began in 1964 with a single set of items that could be used to assess either state or trait anxiety based on rewording the instructions (Form A). The state instructions requested clients to complete items for how they felt “right now, at the moment,” whereas the trait descriptions asked them to indicate how they generally feel. The items were originally derived and adapted from existing anxiety inventories, including the Affect Adjective Checklist (see Spielberger & Reheiser, 2004). Items were reduced and the scale was refined based on the degree to which individual items correlated with the Manifest Anxiety Scale, Anxiety Scale Questionnaire, and Welsh Anxiety Scale of the MMPI (Spielberger et al., 1999). Further evaluation with Form A indicated that merely rewording the instructions was not sufficient to eliminate the clear trait connotations of some of the items. For example, the item “I worry too much” was a good measure of trait anxiety, but merely rewording the instructions indicated that it was not a good measure of state anxiety. As a result, a second form (Form X; Spielberger et al., 1983; also see Spielberger & Reheiser, 2004) was developed based on the trait and state dimensions having their own individual items. Trait items were selected based on their having the highest correlations with the Manifest Anxiety Scale, Anxiety Scale Questionnaire, and Welsh Anxiety Scale, as well as being the most stable over time. The state items were selected based on their being most sensitive to high-versus low-stress conditions (high construct validity) and having the highest internal consistency.

A decade after the publication of Form X, the STAI underwent a further major revision based on factor analysis, a clearer understanding of the concept of anxiety, and an attempt to eliminate item overlap with depression. This resulted in the current (Form Y) version having 20 items for trait and 20 for state anxiety (Spielberger & Sydeman, 1994). Form Y was normed on 1,838 employees of the Federal Aviation Administration, 855 university students, 424 high school students, 1,701 Air Force recruits, and 263 naval recruits. Older persons and those with more education scored somewhat lower than those who were either younger or less educated, which suggests it might be important to use age and education-related norms. Additional norms are available for a neuropsychiatric population, general medical/surgical patients, and young prison inmates. A children’s form, the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973), is also available.

Reliability and Validity

Test-retest reliability for college students over 30- and 60-day intervals indicated reasonably good coefficients ranging between .73 and .86 for trait anxiety. In contrast, state anxiety test-retest reliabilities were relatively lower, ranging from .51 for males to

.36 for females (Spielberger et al., 1983). The lower range for state anxiety is expected, given that state anxiety is considered a more changeable construct, susceptible to many transient factors. Given the expected fluctuations for state anxiety, measures of internal consistency would be more appropriate and important to consider. These have resulted in quite high state anxiety median coefficients ranging between .88 and .93, with a similarly high median trait anxiety coefficient ranging from .92 to .94 (Kabacoff, Segal, Hersen, & Van Hasselt, 1997; Spielberger et al., 1983).

The content validity of the STAI trait scale is supported in that five out of a possible eight domains for a *DSM-IV*-based diagnosis of a Generalized Anxiety Disorder were reflected in the items (Okun, Stein, Bauman, & Silver, 1996). Concurrent validity is supported by high correlations with the Manifest Anxiety Scale and the Anxiety Scale Questionnaire, which ranged from .73 to .75 (Spielberger & Reheiser, 2004). These correlations are sufficiently high that the STAI can be considered an alternative measure of trait anxiety. However, the STAI has the advantage of being shorter and less contaminated by measures of depression. Lower and moderate, but still significant, correlations were found between the STAI trait scale and the Worry Scale (.57) and Padua Inventory (.57; Stanley, Beck, & Zebb, 1996).

The construct validity of the STAI is suggested in that psychiatric patients generally have higher scores on trait anxiety than nonpatient groups (Spielberger et al., 1983; Stanley et al., 1996). One exception is that, as expected, patients with personality disorders tended to have lower scores (Spielberger et al., 1983). Kabacoff et al. (1997) also found that patients with anxiety disorders had slightly higher STAI trait scores than patients without anxiety disorders. Despite this support for the convergent and divergent validity of the STAI, Kabacoff et al. were not successful in developing adequate cutoff scores for identifying the presence of an anxiety disorder. This was primarily because of difficulty finding a score that produced both good sensitivity (high identification of true positives) and good specificity (high identification of true negatives).

Construct validity for the validity of the STAI state anxiety scale is supported in that students during in-class exams and military recruits undergoing stressful training procedures had higher scores when compared to scores taken after relaxation procedures or with age-matched controls (Spielberger et al., 1983). Numerous studies have demonstrated that scores on the STAI are sensitive to the impact of a wide variety of interventions (Newham, Westwood, Aplin, & Wittkowski, 2012; Spielberger et al., 1983; Spielberger & Reheiser, 2004). The STAI has had difficulty discriminating patients with anxiety from patients with depression (Kennedy, Schwab, Morris, & Beldia, 2001), though because of the high comorbidity and general distress related to both syndromes, most measures of anxiety face this same problem.

Factor analyses of the STAI have been mixed. According to the STAI scale development expectations, there should ideally be one factor that loads on trait anxiety and another one on state anxiety (Spielberger et al., 1983). In contrast, Bieling, Antony, and Swinson (1998) found a higher-order factor derived from the trait anxiety items they referred to as *negative affect* and two lower-order factors that they concluded were organized around *depression* and anxiety. Thus, the trait items seemed not to be pure measures of anxiety but included measures of negative affect and depression in addition to anxiety. Whereas Spielberger and Sydeman (1994) did attempt to make Form Y more of a pure measure of anxiety than Form X, this seems to have been

only partly successful. The difficulty in developing a pure measure of anxiety underlies the issue, frequently found in other measures of anxiety and depression, that anxiety and depression have overlapping features with correlations typically ranging between .45 and .75 (Lovibund & Lovibund, 1995). The factor structure of the STAI is further complicated in that Kabacoff et al. (1997) found two factors related to whether the items were worded in a positive or negative direction. They concluded that these items were “method factors” unrelated to the constructs of anxiety. An alternative model by Vigneau and Cormier (2008) found a four-factor model based on a combination of state and trait, as well as whether the items were worded as these variables being present or absent (State Anxiety positive items, State Anxiety negative items, Trait Anxiety negative items, Trait Anxiety positive items). In other words, the bipolar aspects (wording of the items) of trait/state anxiety emerged as independent factors.

Use with Diverse Groups

The utility of the STAI has been demonstrated in a number of cross-cultural contexts. For example, Hishinuma et al. (2000) evaluated the STAI with an Asian Pacific adolescent population and found good internal consistency and a factor structure that was similar to the normative group. Similarly, a Spanish translation of the STAI has been found to have good internal consistency and a comparable factor structure, suggesting it can be used with Spanish-speaking populations (Novy, Smith, Rogers, & Rowzee, 1995). Overall, the STAI has excellent adaptability across a wide variety of ethnic and cross-cultural groups (Spielberger, Moscoso, & Brunner, 2004; Vigneau & Cormier, 2008).

The STAI has also been found to have good psychometric properties with older adult populations (Bergua et al., 2012; Kabacoff, Segal, Hersen, & Van Hasselt, 1997; K. Kvaal, Ulstein, Nordhus, & Engedal, 2005). It has been normed for adolescents and has been found to have good psychometric properties with this population (Spielberger, 1983). Since both adolescent and older adult populations have been found to be somewhat less anxious than adults, separate norms should be used. As noted previously, a version for children between the ages of 9 and 12 is available (STAI-C).

Interpretation

Because the STAI includes two unidimensional subscales, interpretation is mainly composed of descriptions for the variables being measured. Interpretation should consider both the variable itself and the relative magnitude of the person’s score.

High T-Anxiety

The person is likely to perceive a wide number of situations as threatening or dangerous; the person is especially likely to be concerned with being evaluated by other people, with corresponding threats to self-esteem.

High S-Anxiety

The person has feelings of apprehension, worry, nervousness; unpleasant, consciously perceived feelings of tension; the person is also likely to report corresponding activation of the autonomic nervous system.

High S-Anxiety/Low T-Anxiety

The anxiety the person is reporting is likely to be caused by some external threat or a current situational stressor. As a result, it is likely to resolve itself. If intervention is warranted, it should ideally be directed toward problem-solving strategies to help alleviate the current causes of stress. Additional focus may be on strategies that provide a reduction in arousal, such as increasing the person's social supports, systematic desensitization, providing reassurance, hypnosis, exercise, meditation, or progressive muscle relaxation. There might also be an emphasis on what the current anxiety has taught the person about him- or herself and how this might be used to reduce the likelihood of reducing anxiety in the future.

High T-Anxiety/Low S-Anxiety

Although the person is not currently reporting anxiety, he or she is prone to reacting to situations in such a way as to easily become anxious. He or she is likely to be extremely concerned with threats to self-esteem and, as a result, might be apprehensive in any interpersonal situation in which he or she might be judged.

RECOMMENDED READING

- Antony, M. M., & Barlow, D. (Eds.). (2011). *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders* (2nd ed.). New York, NY: Guilford Press.
- Groth-Marnat, G. (1999). Financial efficacy of clinical assessment: Rational guidelines and issues for future research. *Journal of Clinical Psychology*, 55, 813–824.
- Maruish, M. E. (2004). *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed.). Mahwah, NJ: Erlbaum.
- Maruish, M. E., & Nelson, E. A. (2014). *Psychological testing in the age of managed behavioral health care*. Mahwah, NJ: Erlbaum.

TREATMENT PLANNING AND CLINICAL DECISION MAKING

The ultimate goal of psychological assessment is to help solve problems by providing information and recommendations relevant to making the optimum decisions related to the client. Doing this involves integrating a wide variety of information, including specifics of the problem, client resources, a client's personal characteristics, and environmental circumstances. Practitioners must then work with this information to make recommendations related to treatment setting (inpatient/day treatment/outpatient/none), intensity (frequency and duration), goals, mode (individual, group, family), and specific strategies and techniques. The sheer number of these variables can make assessment a daunting task. Thus, the focus of this chapter is to provide a manageable framework for organizing assessment results for planning treatment.

The following format for organizing results and developing treatment plans has been guided by several principles and values. When possible, evidence-based information has been provided. This is possible using the knowledge derived from the rather extensive body of research currently available. In fact, treatment that ignores the procedures indicated by current research runs the risk of not offering clients the most effective treatments available. At the same time, it is acknowledged that clinical experience and judgment inevitably need to interact with the research, assessment results, and the uniqueness of the client to generate the best treatment plan. A further guiding principle underlying this chapter is that the format is both sequential and systematic. It is sequential in that, typically, a series of decisions confronts clinicians, beginning with areas such as how restrictive interventions should be and ending with issues such as specific techniques of therapy and methods of relapse prevention. Finally, the number of variables considered has been reduced to those that seem most relevant, easily manageable, and best supported by research.

Developing effective recommendations requires a number of knowledge and skill areas beyond merely test interpretation. One of the more important areas relates to general case management. Effective case management requires practitioners to survey the general case issues, focus on the most salient features, and make recommendations accordingly. Recommendations should include noting how restrictive treatment should be, which is directly related to the severity of the problem and whether the patient is likely to present a danger to self or others. After reviewing these considerations, practitioners need to be aware of the resources available in the community and make recommendations to the most appropriate one(s), such as a specific inpatient setting or referral to an outpatient clinic, medical facility, suicide prevention center, Alcoholics Anonymous, or behavioral medicine unit. A variety of self-help resources may also help to enhance treatment. Decisions need to relate to the frequency and duration of

treatment. Practitioners should also be able to assess and provide recommendations on how to optimize a client's environment. For example, assessing the client's level of social support might help either in encouraging the person to use available supports or in enhancing only partially adequate supports. Environments might also be changed to increase social interaction or decrease the likelihood of relapse.

Practitioners can and should be able to deliberately tailor their responses toward specific characteristics and circumstances of the client. While this might seem self-evident, many therapists typically provide the same or at least similar interventions for all their clients. Frequently, these interventions are based on the specific school of therapy the therapist is most familiar with (e.g., cognitive therapy for every client who comes in for treatment). Research has demonstrated, however, that whereas cognitive behavioral therapy (CBT) can be effective for patients with externalizing coping styles, a supportive, self-directive method is more effective for patients with internalizing styles of coping (Beutler, Clarkin, & Bongar, 2000; Harwood, Beutler, & Groth-Marnat, 2011). A further assumption frequently found in clinical lore is that empathy is an essential ingredient of all effective therapy. Despite this assumption, controlled studies indicate suspicious clients with low motivation do poorly when psychotherapists are empathic, involved, and accepting (Beutler, Crago, & Arizmendi, 1986; Beutler, Harwood, Alimohamed, & Malik, 2002). (Despite these findings, overall, empathy does seem to have a general positive effect on psychotherapy; see Elliott, Bohart, Watson, and Greenberg, 2011.) These examples, and many others, indicate that not all treatments and treatment styles are optimal for all clients. Recommendations and interventions should, as much as possible, be guided by research, as clinical lore can sometimes be misleading.

This brief introduction to treatment planning is not intended to minimize either the tremendous impact that the quality of the treatment relationship has on outcome or of the importance of clinical experience. The overall quality of the therapeutic relationship accounts for at least as much of the outcome variance as specific techniques (Lambert & Barley, 2002; Norcross, 2011; Wampold & Imel, 2015). Well-defined techniques, however, are often easier to specify and control than the more general quality of the relationship. In addition, techniques that match a client's needs and expectations are likely to enhance the quality of the working relationship. Thus, it is difficult, if not impossible, to separate technique and relationship. For example, relationship quality is likely to deteriorate if a therapist tries highly directive techniques with quite defensive clients (Beutler, Moleiro, & Talebi, 2002; Beutler, Sandowicz, Fisher, & Albanese, 1996). In addition, clinical experience will always be crucial in integrating a diverse range of client information into an optimum set of recommendations. While this process should be generally guided by available research, the specifics of a particular case might be sufficient to alter or even negate the generalities suggested by research data alone. Thus, research findings and clinical information should ideally be in an active interplay such that they optimize each other's strengths and minimize their respective weaknesses.

DEVELOPMENT AND APPROACHES TO TREATMENT PLANNING

One of the central concerns for researchers and clinicians refining treatment planning has been efforts to understand how and why therapeutic interventions do or do not

work. Similar to the debates on intelligence, researchers and clinicians can be divided into “splitters,” who have focused on the impacts of specific techniques, and “lumpers,” who have been more concerned with the common, nonspecific ingredients that facilitate change. A further related theme is the identification of relevant client domains or behaviors needing change and matching these with appropriate interventions. The general purpose of assessment in this process is to identify the most relevant client characteristics or symptom behaviors and match these with optimal interventions. Gordon Paul (1967) ambitiously stated this agenda with a question: “*What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?*” (p. 44).

Ancient traditions of mental health were fully aware of the importance of tailoring interventions toward the specifics of the client. For example, the Vedas discuss the differential effects of telling appropriate metaphors to clients according to their needs. Similarly, Sufism has had a well-developed tradition of storytelling designed to create specific impacts on listeners (Groth-Marnat, 1992). As early as 1919, Freud was concerned with matching patients to different types of psychotherapy. Classical psychoanalysis was recommended for patients who were quite psychologically minded. In contrast, clients who were considered “unanalyzable” because of a lack of psychological sophistication were referred for psychoanalytic psychotherapy, which focused on direct suggestion rather than extensive insight and in-depth self-exploration.

Throughout the 1950s and 1960s, an extremely diverse number of therapies were developed. Each one provided a different theoretical model for causation and a wide variety of techniques. Part of what stimulated these developments was the hope that a series of techniques would prove successful in treating certain types of problems. Examples of such techniques included systematic desensitization for phobias and interpreting the transference as a tool in resolving past interpersonal conflicts. In the psychosomatic literature, it was believed that certain disorders (e.g., asthma) were the result of specific types of conflicts (e.g., suppressed dependency needs). Resolving these specific conflicts, it was hoped, would similarly remove the relevant symptoms. This extensive variety and specificity has led to the development of more than 400 different types of psychotherapies, only a few of which have been subjected to any degree of empirical investigation.

Psychological assessment during the 1950s and 1960s closely paralleled the particular school of therapy it was aligned with. Because many assessment procedures were both used in a medical context and relied on projective techniques, they, accordingly, reflected a psychoanalytic perspective. The goal, then, was to list a patient’s symptoms along with a dynamic interpretation of the conflicts believed to be causing these symptoms. The specificity of treatment planning was deemphasized in favor of detailed descriptions of inner dynamics. It was assumed that, by describing these conflicts, the therapist would know better how to proceed. During the 1960s and 1970s, the competing schools of behaviorism and humanism developed their own modes of assessment based on either specifying target behaviors and the antecedent events leading to these behaviors or attending to the ongoing experience of the client. In either case, the value of traditional psychometric procedures was not only deemphasized but even criticized and abandoned.

Understandably, there was considerable competition between the different therapies as to which one was most effective. In 1952, Eysenck stimulated considerable

controversy with his verdict that psychotherapy (particularly psychoanalysis) was no more effective than placebo. In contrast, he concluded that behavior therapy has demonstrated positive outcomes beyond merely placebo effects (Eysenck, 1994). Much of the ensuing research became a horse race in which proponents of particular schools wanted to demonstrate the superiority of the chosen therapeutic mode that they had received training in for so many years. The classic and much-cited summary study of therapeutic outcome was M. L. Smith, Glass, and Miller's (1980) meta-analysis, which concluded that all of the evaluated therapies were effective. Smith et al. also found greater effect sizes for those therapies with a narrow focus than for those with a wider focus. For example, techniques such as systematic desensitization and hypnosis, which typically target a narrow band of behavior (elimination of a phobia, habit modification), were found to have greater impact than client-centered therapy, with its more general goal of personal growth. However, the differences between the various therapies were not extensive, which led many reviewers of the field to agree with Luborsky, Singer, and Luborsky's (1975) earlier verdict that "Everybody has won and all must have prizes" (often referred to as the "dodo bird" verdict). The "dodo bird" verdict is supported by more recent, methodically well-designed studies that have demonstrated little or no differential outcomes between different therapies when targeted at the same problems (Ahn & Wampold, 2001; E. Anderson & Lambert, 1995; Patterson, 1989; M. E. P. Seligman, 1995; Wampold & Imel, 2015). For example, current high-quality research (including a meta-analysis) has found that randomly assigned, manualized cognitive behavioral versus psychodynamic-interpersonal interventions for depression had similar effectiveness on therapeutic outcome (Gallagher-Thompson & Steffen, 1994; Leichsenring, 2001; D. A. Shapiro et al., 1994).

The preceding studies, along with responses to these findings, have significant implications for treatment planning. One category of response is an investigation of the *nonspecific features of therapy* common to all systems (see Ahn & Wampold, 2001; Andrews, 2001; J. D. Frank & Frank, 2004; Lambert & Barley, 2002; Wampold & Imel, 2015). Underlying this response is the hope that these nonspecific factors would explain the general equivalence of outcomes across therapies. The earliest formal conceptualization was a 1957 description of "necessary and sufficient conditions of therapeutic change" by C. Rogers (1957/1992). These conditions included genuineness, unconditional positive regard, and accurate empathy. A somewhat similar nonspecific formulation was also proposed by J. D. Frank (1973), who emphasized that successful therapy involved providing the client with hope, overcoming demoralization, and creating a corrective emotional experience involving benevolent persuasion. This nonspecific focus provides a contrast to the more directive, technique-oriented approaches. In particular, the nonspecific explanations place considerable emphasis on the quality of the therapeutic relationship beyond mere technique (see Norcross, 2011). The implications for assessment and treatment planning are that assessment results should help make decisions that will enhance the quality of the therapeutic relationship (Andrews, 2001; Luborsky, 1994; Wampold & Imel, 2015). In addition, the specificity of treatment recommendations is also deemphasized. What still remains, however, are basic case management issues (restrictiveness, format, and intensity of treatment) and enhancing aspects of the relationship that are likely to maximize outcome (e.g., matching client expectations, being perceived as trustworthy and credible).

A second general strategy has collectively been referred to as *differential therapeutics*. This approach focuses on refining intervention techniques based on specific diagnoses combined with additional information related to aspects of the problem (see Antony & Barlow, 2011; Nathan & Gorman, 2015; Sammons & Schmidt, 2001). The general function of assessment in differential therapeutics is to diagnose and evaluate the specifics of a disorder as carefully as possible. Techniques believed to be most effective in optimizing outcome are tailored and directed toward a symptom or symptom cluster. This model closely parallels and draws on procedures used in medicine, which similarly rely on accurate diagnosis before applying the optimal treatment.

The preceding approach has had varying degrees of success. Probably the most noteworthy of these successes has been the development of specific targeted interventions for clusters of anxiety-related symptoms (Barlow, 1988; Barlow, Conklin, & Bentley, 2015; J. G. Beck & Zebb, 1994; Steketee, 1994). In particular, Barlow, Craske, Cerny, and Klosko (1989) developed a specific targeted treatment for panic disorder that has been found to be effective for 80% to 100% of those who completed the program. In addition, it has been found to provide outcomes clearly superior to pharmacotherapy (Gould, Otto, & Pollack, 1995). Additionally, differentially effective interventions for obsessive-compulsive disorder have primarily centered on gradual exposure to the anxiety-related situations, along with strategies to prevent the occurrence of the compulsive behaviors (Riggs & Foa, 1993).

Although advantages of using interventions targeted directly at the subtype of disorder (diagnosis) has been shown for most of the anxiety disorders, less success has been achieved for specific interventions in the treatment of depression. The extent of vegetative symptoms, presence of manic episodes (bipolar), and presence of suicidal risk have implications for type of medication and restrictiveness of treatment. Although research has so far not been able to clearly identify the best psychosocial intervention for depression (Gallagher-Thompson & Steffen, 1994; Leichsenring, 2001; Rude, 1986; D. A. Shapiro et al., 1994), some have argued for the differential effectiveness of cognitive behavioral approaches (see Antonuccio, Danton, & DeNelsky, 1995). Researchers have also had difficulty demonstrating differential effectiveness for specific psychosocial interventions for schizophrenia, sleep disorders, sexual disturbances, generalized anxiety disorder, and personality disorders (Beutler & Crago, 1986; T. Brown, O'Leary, & Barlow, 1993).

A third general response has been to consider the nonequivalence of therapeutic outcomes to be the result of insufficiently explored *client characteristics* (see Beutler et al., 2000; Harwood et al., 2011; Groth-Marnat et al., 2008). Evaluating the relative contribution of client characteristics is based on the finding that some types of clients do quite well when provided with a certain type of therapy, while others, given the same therapy, do quite poorly. If those clients who did poorly could have been identified and provided with different strategies, they might have made significant therapeutic gains using an alternate approach. However, the averaged scores on outcome studies using heterogeneous populations have obscured these potentially relevant client differences.

The strategy, then, has been to thoroughly research a wide variety of client characteristics to determine which ones can be used to predict differential response to therapy. Over 200 of these characteristics have been suggested, of which 100 have been subjected to empirical investigation (Garfield, 1994; Harwood et al., 2011; Norcross, 1997).

The result has been that, over the past 20 years, there has been increasing delineation and use of the most empirically validated characteristics for systematic treatment planning (Beutler, Forrester, Gallagher-Thompson, Thompson, & Tomlins, 2012; Beutler, Forrester, Holt, & Stein, 2013; Beutler, Harwood, Bertoni, & Thomann, 2006). Reviews of this strategy have indicated that, under optimal matching conditions, up to 64% of the outcome variance can be accounted for (Berzins, 1977; Beutler, 1983, 1989; Beutler et al., 2012). When client characteristics and treatment matching are combined with the quality of the therapeutic alliance, prediction of outcome increases to 90% (Beutler et al., 1999; Beutler, Moleiro, Malik, & Harwood, 2003). In contrast, providing therapeutic techniques without considering predisposing client characteristics has been found to account for only 10% of the outcome variance (Beutler, 1989; Wampold & Imel, 2015).

The implication for assessment is that predisposing client characteristics can and should be used to identify relevant dimensions. Furthermore, these dimensions should then be used to develop optimum treatment plans. This finding emphasizes both the technical and clinical aspects of assessment, as well as the specificity of treatment recommendations. It does not negate the importance of common factors (caring, empathy, respect, etc.), but selecting treatment based on client characteristics can potentially add to the effects of these common factors. It can also improve outcomes beyond merely assigning patients to models of therapy (e.g., cognitive therapy; Beutler et al., 2003).

In addition to the preceding three general strategies, a variety of specific attempts has emerged to provide guidelines for prescriptive matching of client characteristics with therapeutic interventions. Ideally, the *DSM* should be useful in developing treatment plans in a similar manner as occurs for specific disease entities in general medicine. Generally, however, this has not been the case. Although some of the diagnostic categories have implications for different forms of medical interventions (e.g., antidepressants for depressive disorders), they generally are not particularly helpful for designing psychosocial interventions (Beutler, 1989; Houts, 2002). In an effort to more clearly identify the full array of relevant domains for intervention, Lazarus (1973, 1989, 2005) suggested that clinicians analyze a patient's behaviors, affects, sensory experiences, imagery, cognitions, interpersonal relationships, and need for drugs (BASIC-ID). A somewhat different perspective has been taken by authors who believe that the various stages of therapy or change are crucial to consider in tailoring interventions. Prochaska and DiClemente (1984, 2005), for example, encouraged practitioners to tailor their interventions around the stages of precontemplation, contemplation, preparation, action, and maintenance.

In a behavioral medicine context, Wickramasekera (1995a, 1995b) developed a high-risk model for identifying and assessing clients likely to have somatizing complaints. The high-risk model makes predictions based on accounting for predisposing factors consisting of either very high or very low hypnotizability, neuroticism (level of sympathetic reactivity), and catastrophizing cognitions. Precipitating factors relate to major life changes or minor hassles, and client factors that are likely to serve as buffers include level of social support and coping ability. Treatment can then be tailored toward the patterns of scores on these client dimensions.

A further strategy has been to determine the factors involved in creating optimal matches between therapist and client. In some ways, similarity between client and

therapist has been found to be advantageous, particularly for dimensions such as age, gender, and ethnicity (Beutler & Clarkin, 1990). Similarity is also likely to enhance the value placed on interpersonal treatment goals, friendship, and social recognition (Arizmendi, Beutler, Shanfield, Crago, & Hagaman, 1985; Talley, Strupp, & Morey, 1990). However, research has shown that similarity, including ethnic matching of the therapist-client dyad, does not in itself improve therapeutic outcomes; rather, acknowledging client preference, which may or may not be for an ethnically matched therapist, does impact outcomes (Cabral & Smith, 2011). In contrast, dissimilarity between client and therapist predicted better outcomes when therapists who valued a high level of autonomy worked with clients who had a high need for attachment and dependence. Conversely, therapists who were highly oriented toward attachment and dependency did better with clients who were highly self-sufficient and autonomous (N. S. Jacobson, Follette, & Pagel, 1986).

The relevance and urgency of working with evidence-based methods of treatment planning are likely to increase significantly in the future. A powerful factor fueling this urgency is the current managed care environment, which increasingly demands that both assessment procedures and interventions demonstrate their cost-effectiveness (Groth-Marnat, 1999; Groth-Marnat & Edkins, 1996; Groth-Marnat et al., 1995; Maruish & Nelson, 2014; Yates & Taub, 2003). As a result, there is increasing pressure to demonstrate that assessment can quickly identify client problems, facilitate optimal treatment recommendations, and show the effectiveness of actual interventions. These “tools of the trade” must be able to provide these services in a way that has been demonstrated to be cost-effective. At the present time, data on the cost-effectiveness of assessment are not yet available, but it is hoped they will be forthcoming in the near future (Groth-Marnat, 1999, 2000b; Yates & Taub, 2003). Future research should clarify when assessment is and is not cost-effective and, in particular, demonstrate that assessment results can be used to save money by quickly and effectively developing a treatment plan, thereby avoiding misplaced and possibly ineffective or unnecessarily long treatment.

INTERVENTION OPTIONS

Groth-Marnat and Davis (2014) identified six different categories of potential recommendations that could emerge from a psychological assessment:

1. Treatment
2. Placement
3. Further evaluation
4. Alteration of the client’s environment
5. Education and self-help
6. Other miscellaneous recommendations

Treatment, which can include therapy, medication, and other interventions that can be provided by other professionals, is the focus of much of the rest of this chapter.

Placement recommendations include decisions related to where a client will receive the least restrictive necessary services. The overwhelming majority of clients will benefit most from some sort of outpatient treatment placement, such as at a community mental health clinic or seeing a therapist in private practice. However, some decisions need to be made about the need for a higher level of care. Information to evaluate when deciding if an individual would most benefit from and needs inpatient hospitalization includes severity of symptoms and ability to self-care (George, Durbin, Sheldon, & Goering, 2002), as well as danger to self, severity of psychosis or depression, and level of impulse control (Way & Banks, 2001). Day treatment and the more restrictive residential treatment are options for those who require a higher level of care than outpatient treatment but do not warrant inpatient hospitalization. Walton and Elliott (1980) identify several factors to take into consideration when deciding if residential treatment is necessary. These include the psychological, physical, and social problems of the individual; the actual or potential for problematic behavior; and the community's tolerance for that actual or potential problematic behavior. More specifically, Chor, McClelland, Weiner, Jordan, and Lyons (2012, 2013, 2015) have recommended residential treatment for individuals (they write mostly about youth) who have complex (multiple) emotional or behavioral problems that are both dangerous and addressable in some form of treatment. Other decisions that need to be made, other than outpatient versus day treatment, residential, or inpatient, include voluntary versus involuntary options (based on the level of need and the level of resistance). Additionally, legal guardianship or supported decision making (see Gooding, 2013) may be recommended if an individual does not have the capacity to make adequate decisions for him- or herself.

Recommending *further evaluation* is most appropriate when, during the course of an assessment, the evaluator decides that there is some information that may be pertinent to the client and his or her current problems that are outside of the area of expertise or competence of a psychological assessor. Perhaps most common is the need for medical consultation or testing, such as testing for thyroid dysfunction, neurological damage, or some other medical or physical factors that are known to significantly impact psychological and psychosocial functioning. Other examples include legal or financial problems that are revealed during the assessment that need further evaluation with an expert in those fields, evaluation for the potential benefit of psychotropic medication, or (most often with children) further evaluation by a speech or occupational therapist. It is important for those conducting psychological assessment to fully understand the limits of their competence and to refer, as appropriate, for further evaluation when needed. In some cases, this may even mean referring to another psychologist. One common example might be a referral to a neuropsychologist when a personality assessment reveals the possibility of neuropsychological dysfunction, but the assessing psychologist is not a neuropsychologist and is not competent to fully evaluate neuropsychological functioning.

Groth-Marnat and Davis (2014) discussed a class of recommendations related to *altering the environment* of the client in some way. This strategy has been utilized extensively in neuropsychological contexts, such as recommending reminders to take medication, installing stove burners that shut off automatically after a certain amount of time, or setting timers with reminders for certain tasks. Additionally, in psychoeducational settings, evaluators may recommend altering the learning environment in certain

ways, such as changing the seat of a child in the classroom in order to reduce distraction; adding a note-taker or paraprofessional for support; or altering the contextual factors for taking tests, such as allowing a laptop, extra time, or providing a quiet room away from the normal classroom in which to take them. Other environmental alterations may be appropriate as well, such as ridding the home of alcohol (including asking a spouse not to drink it at home) to prevent relapse, placing electronics outside of the bedroom at night, or assigning a friend or family member to regularly check in on the client in certain ways.

Self-help has become a large industry over the past 40 (and even more) years, and many programs (books, computer-assisted, etc.) include psychoeducational and cognitive behavioral strategies. Norcross et al. (2013) provided a comprehensive guide to evidence-based self-help resources, and I. M. Marks, Cavanagh, and Gega (2007) presented resources for computer-assisted psychotherapy. Self-help, if well implemented, can have extremely good effect, almost with the same effect size as professional counseling (den Boer, Wiersma, & van den Bosch, 2004). Effective implementation of self-help resources is extremely important, and an assessment can include in its recommendations specific tips for implementation, especially if the self-help resources will be the primary mode of treatment, without the ongoing assistance of a mental health professional. Norcross (2006) offered suggestions about effective implementation of self-help resources. In addition to books, movies, Internet material, and computer-assisted treatment, self-help and support groups can also be important recommendations. Assessors should familiarize themselves with local resources before recommending support groups, to ensure that clients have access to them.

Finally, *miscellaneous recommendations* are recommendations that do not fit neatly into the above categories. Examples may include recommending that an individual surrender his or her driver's license or recommending potential educational or career options to pursue. Additionally, it is often useful to articulate clearly with whom the assessor recommends the client or referral source shares the report. An alternative to this is recommending that the client or referral source sign a release of information consent form so that the evaluator can give feedback directly to others who may benefit from hearing it (such as a treating therapist, a teacher, potential supports, etc.). Finally, practitioners should always remember that it is a potentially valid decision to recommend no treatment at all, if the assessment reveals no need for it.

CLINICAL DECISION MAKING

When creating recommendations within the context of an assessment report, two major considerations must be taken into account. First, and the primary focus of this chapter, is the selection of appropriate, hopefully effective recommendations that will help improve the individual's life in some way. The following sections of this chapter help guide the decision-making process related to selecting effective treatments and other recommendations. Although research is helpful, recommendations should be tailored specifically to the client (Groth-Marnat & Davis, 2014). This means taking into account many different client factors, many of them addressed in this chapter, but some of which will be unique in each individual case and will thus require clinical skill

and knowledge gained through the relationship built during the assessment. Much of the data unique to the individual being assessed are organized into a clear, coherent formulation of the case. Second, and also important, is communicating the recommendations in a way that is likely to be understood by the client and in a way that increases the likelihood that he or she will follow through on them. In order to accomplish this, the wording of the recommendations should be tied clearly to the individual who was assessed. This should include cultural considerations (see T. B. Smith, Domenech Rodriguez, & Bernal, 2011), as well as linking each recommendation back to a specific problem or goal that emerged from the assessment. This method also ensures that recommendations are logical, address the referral questions and problems that emerged during the assessment, and are not of greater intensity than needed.

When deciding what interventions to recommend, four overarching areas of consideration should drive clinical decisions. First, the clinician should organize the assessment results into a model that explains, in a narrative way, what is going on with the client. This case formulation often proposes causal mechanisms and provides information on where interventions should be aimed. Second, the clinician should use assessment results to understand the exact problems, in order to inform treatment targets and goals, both short and long term. Factors related to the problems themselves, which can influence the goals for treatment, are the ultimate diagnosis (if a diagnosis is arrived at through an assessment), the complexity (including the chronicity) of the problem, how functionally impaired the individual is (severity), and how much subjective distress the individual is experiencing (which can have great impact on level of motivation). Third, the clinician should consider the context of the problems, which may have contributed to establishing them and, more importantly, may reinforce and maintain them. These contextual factors include an understanding of how the client copes with problems and how effective this coping is, an evaluation of the level and type of social support, and an understanding of other life circumstances that may reinforce the client's problems. Fourth, and finally, clinicians should consider process and context variables that are specific to therapeutic treatment, such as the client's level of resistance, his or her stage of change (see Prochaska & DiClemente, 2005), and other client characteristics that may affect treatment and the treatment relationship (see Norcross, 2011). The sections that follow in this chapter focus on these four overarching considerations for making clinical decisions and developing recommendations as a result of assessments.

CASE FORMULATION

Practitioners typically (though not always) put the assessment data (including test data and other information) together to present a comprehensive narrative of what is going on for the client, within his or her context. Doing this is perhaps the most difficult and nuanced part of psychological assessment. That narrative serves to help the audience of the report and recipients of feedback better understand what is likely going on for the client, beyond just a straightforward description of symptoms. The narrative is generally theoretically informed, which means it is highly dependent on the theoretical orientation of the clinician (e.g., Persons, 2008). That is, the narrative

description of what is likely causing and maintaining the problems of the individual who was assessed relies heavily on what the practitioner believes maintains problems; for example, one clinician may believe that cognitive or behavioral reinforcers maintain the anxiety experienced by an individual, whereas another may believe that underlying dynamics related to repeated interpersonal and attachment patterns serve that maintenance role. Regardless of the clinician's theoretical orientation, most case conceptualizations include at least four components (Eells, Kendjelic, & Lucas, 1998): (1) a description of the symptoms or problems being experienced by the individual; (2) predisposing vulnerabilities of the individual, including life events and internal predispositions; (3) stressors, events, and environmental contextual information that led to the problems; and (4) some sort of hypothesized causal mechanism that links all of these components together to explain why and how the problems are maintained. There are multiple resources for integrating assessment data to provide case formulations (e.g., Beutler, Malik, Talebi, Fleming, & Moleiro, 2004; Butcher & Perry, 2008; Eells, 2011; Harwood et al., 2011; Tarrrier & Johnson, 2006; Wright, 2010). Each one uses different theoretical and practical frames to conceptualize causal mechanisms that sustain client problems, and each one requires knowledge (some more, some less) of psychological theories and mechanisms. The transtheoretical models presented in Wright's (2010) *Conducting Psychological Assessment: A Guide for Practitioners* are discussed here, though practitioners are encouraged to consult the original text for a more in-depth discussion with clear examples.

Wright provided four potential models for organizing data into a case formulation. It is important to note that these are four options of many, and they are compatible with most theoretical orientations. The first, perhaps most widely used model for case formulation is a *Diathesis-Stress Model*, which categorizes assessment findings into three categories: (1) diatheses, which are the factors that an individual him- or herself contributes to the situation, such as personality style, general approach to the world, or vulnerabilities; (2) stressors, including what is happening or has happened within the client's life or environmental context; and (3) outcomes, which are the problems or symptoms themselves. The psychological model is that the diatheses interact with the environmental stressors in order to cause and maintain the outcome problems or symptoms. This model can work with cognitive behavioral theoretical orientations; for example, a schema of low self-worth (diathesis) combined with getting fired from a job (stressor) could reinforce low mood and depressive cognitions (outcomes). The model can also work with a more psychodynamic orientation; for example, a problematic attachment pattern (diathesis) combined with a failed relationship (stressor) may reinforce self-destructive tendencies (outcome). Whatever theoretical orientation the clinician holds, some model within it likely focuses on internal and external components that causally reinforce problems.

The second model presented by Wright (2010) is a *Developmental Model*, in which there is a developmental mismatch between the client's functioning and the demands of his or her life and environment, and this mismatch reinforces problematic outcomes. This model requires the practitioner to have knowledge of normative developmental trajectories and to clearly translate these trajectories into useful feedback descriptions. For example, a young adult who has spent much of his or her life using drugs may function emotionally and interpersonally more like an adolescent, never having acquired

the appropriate skills that most develop during actual adolescence. Because this individual is a young adult, though, the demands placed on him or her are those generally placed on young adults. Because there is a mismatch between the life demands and the social and emotional capabilities of the individual, he or she will likely evidence some problems, such as interpersonal conflict or job difficulty. Again, this model can accommodate multiple theoretical orientations. What is considered normatively appropriate development may be framed in cognitive terms, such as capacity for abstract thought or tendency to accommodate attitudes versus assimilate conflicting information into previously held beliefs; in psychodynamic terms, such as employing healthier and more mature defense mechanisms; or in other terms. The focus is on the mismatch between abilities/internal functioning and what normatively should have been achieved by this age in order to meet the demands of life and the environment.

The third model is the *Common Function Model*, which posits that each of the themes that emerged in the assessment seem to be serving a common purpose or function. For example, multiple problems and processes may emerge that all seem to be serving the purpose of distancing an individual from others. These problems may include a tendency to act out interpersonally, a discomfort with emotional closeness with others, social anxiety, and beliefs that he or she is not worthy of relationships. Each of these factors seem to be serving the same function of distancing the individual from other people, with this underlying desire or process serving as the mechanism for maintaining each one (i.e., because they work, effectively distancing the individual from others, each is reinforced). Again, this model can accommodate different theoretical orientations, depending on what the practitioner believes serves different “purposes” for individuals. Themes may focus on cognitions and behaviors, underlying defensive dynamics, interpersonal strategies and patterns, or a combination of these and more.

The final model presented by Wright (2010) is more flexible than the other three, but it is also more sophisticated and complex. As such, it is referred to as a *Complex Model* for case formulation, and it often reflects the fact that clients are more complex than many prescribed models of functioning can account for. These complex models are often hybrids of the preceding models, with the understanding that some underlying processes may represent multiple factors (e.g., an outcome within the diathesis-stress model may also serve as a reinforcing diathesis for another problem). For example, acting out behavior at school may be an outcome of a child who has an externalizing coping style (see the “Coping Style” section later in the chapter) as a diathesis and whose parents are going through a divorce, which serves as a stressor. The acting-out behavior, the outcome in this diathesis-stressor model, may also serve as a reinforcing diathesis (along with the stressor of parents divorcing) for academic difficulties. In this case, the lack of parental academic support (because of their preoccupation with their own problems during the divorce), combined with behavioral difficulties in school, causes and maintains poor academic performance. This example of a complex model uses one of the previously presented models as a base but adds nuance and complexity to it by extending beyond the structure of the base model. These complex models, then, can also accommodate multiple theoretical orientations, again based on the reinforcing mechanisms that align with the orientations.

The case formulation developed most often has direct bearing on clinical decisions that are made in terms of treatment recommendations. Without a case formulation,

the most logical treatment recommendations would be direct, specific interventions aimed at alleviating the symptoms that emerged during the assessment. However, with causal hypotheses, treatment recommendations can target not just the outcome problems but the proposed underlying mechanisms that sustain them. Depending on the theoretical orientation, these mechanisms may include underlying schemas, attachment patterns and defensive styles, coping skills, environmental risk factors, or other factors that cause and maintain problems. The underlying assumption here is that there are, in fact, underlying dynamics that cause and reinforce exhibited problems and symptoms and that addressing these underlying factors is likely to ameliorate the problems. A focus on underlying mechanisms does not preclude treatment recommendations that directly target outcome symptoms, though. For example, a case formulation may be built around factors that cause and reinforce suicidal tendencies, which is conceptualized as an outcome problem; in addition to treatment recommendations for the underlying mechanisms, the practitioner would likely recommend strategies that directly target the problematic outcome symptom of suicidality, and he or she would likely even recommend this first. The treatment recommendations from there may address underlying factors, once the client's safety is assured.

UNDERSTANDING THE PROBLEMS

Diagnosis

In many ways, diagnosis has been the driving force behind clinical decision making in the mental health field for many decades. Perhaps because much of the research (and research funding) in the field goes to evaluating treatments for specific, defined disorders, there are many resources available for assessors to determine optimal treatments based on diagnosis (e.g., American Psychological Association Division 12—Society of Clinical Psychology, 2009; Antony & Barlow, 2011; Chorpita et al., 2011; Fonagy et al., 2014; Nathan & Gorman, 2015; Ollendick & King, 2004; L. Seligman & Reichenberg, 2014). And, in fact, research has found that inaccurate diagnosis can actually detrimentally impact both the process and outcomes of therapy (Jensen-Doss & Weisz, 2008). However, more and more research is focusing less on diagnosis as the primary decision-making feature in treatment selection and more on other factors, such as the multiple client variables included in this chapter (Beutler & Clarkin, 1990; Groth-Marnat, Roberts, & Beutler, 2001; Norcross, 2011). Despite this shift in focus, however, it is important to note that accurate diagnosis (and especially lack of inaccurate diagnosis) serves as a prerequisite for making decisions about appropriate treatments.

Although many question the value and ethics of diagnostic labeling, diagnoses do provide specific value to treatment providers, especially because they predict a specific likely course of symptoms (First & Tasman, 2004). Some diagnoses are likely to have poorer prognoses than others. For example, those with schizoid and antisocial personality disorders have difficulty engaging in productive therapy, although certain other mental health conditions related to these personality types can often be targeted and treated effectively. However, diagnosis should not be revered as

entirely prognostically predictive, as many other individual factors can influence how a disorder progresses (e.g., Hara et al., 2012; Hser, Evans, Grella, Ling, & Anglin, 2015; Koenen, Stellman, Stellman, & Sommer, 2003; Schoevers, Beekman, Deeg, Jonker, & Tilburg, 2003). Developmental factors (especially in children) and cultural factors should be taken into account when considering how diagnosis may influence treatment recommendations (Ridley & Kelly, 2007; Yasui & Dishion, 2007). Despite these considerations, some diagnoses do in fact “prescribe” certain kinds of treatment. For example, a diagnosis of a psychotic disorder like schizophrenia generally suggests that the client should have a psychiatric consultation, as medication is a predominant form of treatment for active psychotic symptoms. Additionally, typical treatments for schizophrenia include psychoeducation for the client and his or her family, skills training, and vocational rehabilitation (Harding & Strauss, 1985; Paul & Lentz, 1977). Because diagnosis is included in many, if not most, psychological assessments, it is important that it be accurate and be taken into account when considering treatment recommendations; nevertheless, factors beyond diagnosis should also be considered (Nelson-Gray, 2003).

Functional Impairment

A pressing problem related to any assessment is an evaluation of the severity of the problem. The core issue is to assess the extent to which the patient’s problem interferes with his or her ability to deal effectively with everyday social, occupational, and intrapersonal requirements. The degree of functional impairment will have a direct relationship to the client’s ability to cope, ego strength, level of insight, and chronicity of symptoms. In many cases, functional impairment relates to the extent to which the client is subjectively distressed. In many instances, however, subjective distress does not relate to the presence of severe problems. Examples include those with antisocial personality disorder, who create suffering for others but do not feel particularly distressed themselves, and those with schizoid personality disorder, who are functioning on the fringes of society but do not feel particularly worried about their marginal status and level of dysfunction. The major distinction is that functional impairment is reflected in objective indicators of impairment. In contrast, subjective distress does not necessarily mean that the person is also impaired based on objective indicators. In general, severity of problems has been linked to negative treatment outcomes (Castonguay & Beutler, 2006), and it is important to evaluate functional impairment across different contextual domains (Brod, Johnston, Able, & Swindle, 2006; K. Green, Worden, Menges, & McCrady, 2008).

There are numerous formal and informal assessment procedures for assessing functional impairment. Beutler and his colleagues (Beutler & Harwood, 2000; Harwood et al., 2011) have summarized the relevant assessment dimensions to include:

- a problem that interferes with the client’s ability to function during the interview;
- poor concentration during assessment tasks;
- distraction by minor events;
- general incapacity to function;

- difficulty interacting with the clinician; and
- multiple impaired areas of performance in the client's daily life.

A mental status evaluation is one structured way of obtaining useful information related to functional impairment.

One of the more useful psychometric indications of functional impairment is the presence of generally elevated scales on the Minnesota Multiphasic Personality Inventory (MMPI-2)MMPI/MMPI-2-RF. Functional impairment is especially likely if elevations are found on scales on the right side of the profile (Paranoia, Schizophrenia, Hypomania on the MMPI-2; Ideas of Persecution, Aberrant Experiences, Hypomanic Activation on the MMPI-2-RF). High Beck Depression Inventory—II (BDI-II) scores (30 or above) also suggest a high level of incapacity. Suicide level should always be assessed if the client is depressed. Specific signs to alert the clinician to suicide risk are relevant critical items on the MMPI-2/MMPI-2-RF or items 2 and 9 on the BDI-II. General elevations on the Millon Clinical Multiaxial Inventory (MCMI-IV) scales also suggest a high level of functional impairment, particularly if elevations occur on the Severe Personality Pathology or Severe Syndrome scales. Although the multiaxial system of the text revision of the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR*; American Psychiatric Association [APA], 2000) provided methods for summarizing information relevant to estimating functional impairment, such as Axis IV Psychosocial and Environmental Problems and Axis V Global Assessment of Functioning, the *DSM-5* (APA, 2013) does not provide a system to do so. However, some inferences can be made based on diagnosis. For example, impairment is likely to be more severe if there are comorbid personality disorder and non-personality disorder diagnoses and if there is the presence of severe disorders in the psychotic domain (e.g., schizophrenia, bipolar disorder).

Several noteworthy instruments covered in previous chapters (e.g., see Chapter 13) can also provide useful indicators of functional impairment. A high number of reported problems (*T* above 63) on the Brief Symptom Inventory (BSI; Derogatis, 1992) suggests high functional impairment, as do high scores (*T* above 55) on the Trait Anxiety Scale of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

When using functional impairment to help clinical decision making, level of functional impairment has implications for these five areas: (1) restrictiveness of treatment, (2) intensity of interventions (duration and frequency), (3) use of medical versus psychosocial interventions, (4) prognosis, and (5) the urgency of achieving initial goals (see Harwood et al., 2011). Low functional impairment suggests that treatment can be in an unrestricted setting (outpatient) and of relatively low frequency and duration. Psychosocial interventions will likely be the predominant form of intervention, and there will be less urgency to rapidly define and achieve specific, symptom-oriented goals.

Severe problems, particularly if the client is suicidal or cannot function in daily activities, may require immediate inpatient care. Examples of diagnoses that may require inpatient care include bipolar mood disorders, psychotic conditions, major depression with suicidal intentions, acute substance abuse requiring detoxification, and some organic conditions that have resulted in significant decompensation. Initial treatment on an inpatient basis might later be reduced to partial hospitalization or day treatment

when the condition has become stabilized. In contrast, outpatient interventions would be appropriate for the vast majority of clients whose problems are of mild to moderate severity (e.g., adjustment reactions, mild to moderate depression) and who have greater resources.

The intensity of treatment (duration and frequency) varies from client to client based primarily on functional impairment. Greater intensity of treatment is generally suggested for clients with these indicators:

- More serious diagnoses (e.g., borderline personality disorder)
- Poor premorbid functioning
- External stress seemingly of minor importance in the development and maintenance of the disorder
- Age between 25 and 50 years
- The expectation that change takes time
- Exploratory and insight-oriented therapeutic framework
- Low levels of social support

In contrast, these indicators suggest short duration of and potentially less frequent interventions:

- An acute disorder (e.g., adjustment disorder, brief psychotic disorder)
- External stress that seems to be of primary causal significance
- Good premorbid level of functioning
- Expectation for change to occur quickly
- Symptom-oriented focus of treatment, or crisis intervention
- Structured, directive, and active interventions
- Children and older adults
- High levels of social support

For some conditions, intermittent brief therapy throughout the life span at critical junctures might be an appropriate recommendation. At times, it might be appropriate to recommend no treatment, particularly if the person might have a negative response (e.g., some individuals with borderline personality disorder or paranoid ideation), no response (e.g., some individuals with antisocial personality disorder), spontaneous improvement (e.g., normal grief), or strong response to suggestions that he or she will improve rapidly with no treatment. Additional characteristics contraindicating psychotherapy might be a client's associating emotional pain with the change process, suspiciousness toward the therapist, and need for control (Mohr, 1995).

Conditions like schizophrenia, bipolar disorder, or severe anxiety states might require medical intervention (pharmacotherapy, electroconvulsive therapy [ECT]) to enable clients to function well enough to become engaged in psychosocial or environmental interventions (see Sammons & Schmidt, 2001). Markers for such interventions might include poor orientation to time and place, poor short-term memory, marked confusion, or clearly inappropriate mood. Past clinical and research evidence

has suggested that severe depression responds better to pharmacotherapy, whereas situationally caused mild and moderate depression responds better to psychosocial interventions (e.g., Elkin et al., 1995). In contrast, evidence indicates that both severe depressions and mild to moderate depression can be treated at least as effectively with psychotherapy but without the potential for problematic side effects (Antonuccio et al., 1995; Casacalenda, Perry, & Looper, 2002; Craighead, Johnson, Carey, & Dunlop, 2015; Free & Oei, 1989; Garvey, Hollon, & DeRubeis, 1994; McLean & Taylor, 1992; Simons & Thase, 1992). It may be that individual characteristics, including level of functional impairment, affect how an individual will differentially respond to different forms of treatment (Craighead & Dunlop, 2014; Sotsky et al., 2006). A clearer indication for antidepressant medication is a high number of vegetative symptoms (e.g., fatigue, insomnia, loss of appetite; Preston, O'Neal, & Talaga, 2005). Similar decision processes can be made for anxiety, psychotic, and bipolar disorders (see Preston et al., 2005).

Functional impairment is one thing to consider when making prognostic judgments, in addition to diagnosis, chronicity, subjective distress, and client resources (employment, abilities, social support). Research on prognosis is somewhat contradictory. On one hand, it might be argued that a person with a severe problem will have difficulty overcoming it because it has progressed to such an extensive level. On the other hand, functional impairment may represent an extreme level in a fluctuating condition so that the person is likely to return spontaneously to an improved level of functioning. In addition, the potential magnitude of change is likely to be greater because the person has so much room for potential improvement. One guideline is that a high degree of psychiatric symptoms associated with the presence of somatic complaints (e.g., headaches, irritable bowel syndrome) is likely to suggest a poor prognosis (Blanchard, Schwarz, Neff, & Gerardi, 1988; Jacob, Turner, Szekely, & Eidelman, 1983). In contrast, patients presenting with severe levels of general anxiety and ambulatory depression typically do quite well with either psychosocial or pharmacological interventions (Elkin et al., 1989).

Finally, severe problems suggest that the urgency of treatment is greater and should be focused around working with the symptomatic areas causing the client the greatest problems and distress. Less severe problems mean that the urgency for change is less, and the goals can change and be negotiated over time.

Problem Complexity and Chronicity

Some clients present with problems that are narrow, focused, and either reinforced by or elicited by the environment, such as simple phobias. In contrast, other clients present problems of a diverse, complex nature. These problems are likely to be pervasive and enduring and occur in many contexts. Instead of focusing on one or two specific behaviors, they involve diverse themes. A review of past relationships typically reveals that these themes have been enacted with persons in intimate relationships or those in positions of authority. Examples might include passive-aggressive interactions with authority figures, conflicts between dependency and independence in intimate relationships, or consistently creating problematic relationships by choosing incompatible partners (e.g., many individuals with alcoholism) despite the availability of more

appropriate persons. These themes can be considered reenactments of internal, unresolved conflicts. For these sorts of problems, the level of intervention needs to be quite different from problems that are narrow and symptomatic. More chronic problems have been linked to worse treatment outcome in general (Castonguay & Beutler, 2006). It is a rule of thumb that the greater the chronicity of the disorder, the more difficult it is to treat.

Problem complexity can be differentiated from functional impairment in several ways. Whereas *functional impairment* refers to level of dysfunction, *problem complexity* refers to underlying thematic patterns in the person's life that may or may not result in a high level of impairment. For example, a client may be functioning at a rather high level (low functional impairment) but still be quite troubled by chronic dissatisfactions in his or her relationships. These dissatisfactions may be the result of complex patterns related to difficulties dealing with anger or issues related to dependency. Such patterns may pervade not only one or two primary relationships but most of the people the person comes into contact with. Whereas severe problems might be quite directly caused and reinforced by the environment (e.g., habits, reactions to stress), a complex problem is likely to be strongly related to internal, unseen events. Furthermore, complex problems are likely to involve personality patterns that are spread across a wide variety of domains.

Problem complexity is more difficult to measure than most of the other factors relevant for treatment planning, in part because it is more theoretically bound. Clinicians from psychodynamic perspectives are far more likely to frame client difficulties as centering around symbolic, underlying, complex themes, whereas behaviorally oriented practitioners describe problems in narrower, concrete, environmentally oriented language (Wittman & Koele, 1999). Although there is no clear resolution to this dilemma, three main features can be used to indicate problem complexity. The first is the presence of several concurrent problem domains or diagnoses (comorbidity), and the second is the presence of pervasive or recurrent patterns and themes of problem behaviors. A third feature suggesting a complex problem is the presence of a personality disorder or, at least, a personality style suggestive of a personality disorder. Beutler and his colleagues (Beutler & Clarkin, 1990; Beutler & Harwood, 2000; Gaw & Beutler, 1995) have summarized indicators of problem complexity based on these background information and behavioral observations:

- Behaviors are repeated as themes across unrelated situations.
- Behaviors are ritualized efforts to resolve underlying interpersonal or dynamic conflicts.
- Interactions seem related primarily to past rather than present relationships.
- Suffering, rather than gratification, is the result of the repetitive behavior.
- Problems are symbolic expressions of underlying unresolved conflicts.

In contrast, noncomplex problems are more often characterized as being:

- Situation-specific;
- Transient;

- Based on inadequate knowledge or skills;
- Having a direct relationship to initiating events; and/or
- Stemming from chronic habits.

Another reason problem complexity is difficult to assess is that there are no clear, well-defined instruments to do so. However, some inferences can be made from existing tests. In particular, elevations on the Millon inventories are likely not only to suggest the presence of a complex problem but also to provide information related to personality themes (Choca & Van Denberg, 2004; Millon & Bloom, 2008). The presence of a personality disorder as defined by *DSM-5* (APA, 2013) criteria further suggests a complex problem. Additional information can be derived from themes noted in Thematic Apperception Test (TAT) story content or from the client's organization of his or her responses to the Rorschach. Both of these instruments can be quite useful in articulating how a client copes with his or her emotions, responds to stress, resolves conflicts, relates interpersonally, and defends against anxiety. Finally, the MMPI-2/MMPI-2-RF can help clarify not only a client's symptom pattern but also the dynamic interplay among the symptoms, coping strategies, likely patterns in interpersonal relationships, and overall personality structure. For example, on the MMPI-2, a chronic problem is indicated if Scales 1 (Hypochondriasis) and 2 (Depression) are both above 65, but Scale 1 is clearly higher (5–10 points or more) than 2. Problem chronicity is also suggested if both Scales 7 (Psychasthenia) and 8 (Schizophrenia) are above 65, but Scale 8 is clearly higher (5–10 points or more) than 7 (see Chapter 7).

Problems that are low in complexity can be treated effectively by targeting specific symptoms, antecedents that elicit these symptoms, and consequences that maintain them. Depending on the problem, specific techniques might include any of those listed next:

- Behavioral contracting
- Social skills training
- Graded exposure
- Reinforcement of target behaviors
- Contingency management
- Challenging dysfunctional cognitions
- Practicing alternative cognitions
- Practicing new self-statements
- Self-monitoring
- Paradoxical strategies
- Counterconditioning
- Relaxation
- Deep muscle relaxation
- Biofeedback

Complex problems, in contrast, are likely to respond best to broad treatments that are directed toward resolving long-standing underlying conflicts and changing patterns

of interpersonal relationships. Depending on the problem, specific techniques might include any of those listed next:

- Group or family therapy exploring patterns of responses
- Cathartic discharge
- Enacting opposite patterns of how the client typically behaves
- Exploring thematic patterns in behavior and relationships
- Interpreting the transference
- Interpreting resistance
- Treatments that address multiple broad issues, such as dialectical behavior therapy and interpersonal therapy

Subjective Distress

Subjective distress relates to the degree to which the person subjectively experiences his or her problems. Such distress is manifested primarily in heightened anxiety, confusion, or depression. A moderate level of subjective distress is useful because it motivates a client to become involved with change. It can lead to cognitive improvements, including enhanced memory, faster performance, and higher intellectual efficiency. If a client's distress becomes too high, however, it will be disruptive and result in deteriorated ability to function. The person then has difficulty appropriately processing information and concentrating, which interferes with the problem solving and behavioral experimentation required in therapy. A client whose level of subjective distress is too low will have difficulty becoming engaged in actively working to change behavior. Thus, there is an optimum window of distress that clinicians should try to achieve (Beutler & Harwood, 2000; Gaw & Beutler, 1995; Harwood et al., 2011).

Although functional impairment and subjective distress have some overlap, there are also a number of differences. As discussed previously, degree of functional impairment relates to objective indicators of poor functioning, whereas subjective distress is more an internal, subjective phenomenon. In addition, subjective distress can be quite changeable and may be controlled by environmental events. A client's level of subjective distress should be monitored from session to session or even within each session. A further contrast exists in the range and types of decisions relevant to either functional impairment or subjective distress. Issues relevant to functional impairment require wide-ranging decisions related to treatment setting (inpatient/outpatient), prognosis, treatment intensity (duration and frequency), and the general goals of intervention. The treatment implications of subjective distress are much narrower in that they provide guidance on whether arousal should be increased or decreased.

Frequent review of interview data, including behavioral observations and relevant history, is one of the best methods of monitoring a client's distress levels. Specific indicators of high distress include (Beutler & Harwood, 2000; Gaw & Beutler, 1995; Harwood et al., 2011) those listed next:

- Motor agitation
- High emotional arousal

- Poor concentration
- Unsteady voice
- Autonomic symptoms
- Hyperventilation
- Hypervigilance
- Excited affect
- Intense feelings

In contrast, low levels of distress are indicated by these indicators:

- Reduced motor activity
- Poor emotional investment in treatment
- Low energy level
- Blunted or constricted affect
- Slow speech
- Unmodulated verbalizations
- Absence of symptoms

MMPI-2/MMPI-2-RF scales that are especially sensitive to subjective distress are: F/F-r, 2 (Depression)/RC2 (Low Positive Emotions), and 7 (Psychasthenia)/RC7 (Dysfunctional Negative Emotions). Collectively, these scales are frequently referred to as the distress scales (see descriptions under “*F* Scale (Infrequency),” Scales 2 and 7, and the 27/72 code type in Chapter 7). However, motivation to change might be undermined if scales related to denial, resistance, and defensiveness are elevated (e.g., *L/L-r* and *K/K-r*, as well as 3-Hysteria/RC3-Cynicism). A poor prognostic sign is a low 7 (Psychasthenia)/RC7 (Dysfunctional Negative Emotions) with elevations on other scales suggesting psychopathology. This finding suggests that the client might be unrealistically relaxed regarding his or her difficulties or has given in to the inevitability of the problems. Additional measures of subjective distress are the Symptom Checklist 90–R, Brief Symptom Inventory (BSI), and the State-Trait Anxiety Inventory (STAI; see Chapter 13). A high level of distress is suggested if the Global Severity Index on the BSI is above 63 or the State Anxiety Score is in the top quartile.

Clients with low subjective distress may be referred involuntarily. Experiential strategies can confront clients with the impact and consequences of their difficulties and are likely to increase distress to a level that makes them more open to changing their behavior. Possible techniques are listed next.

- Two-chair work
- Symptom exaggeration
- Experiential role plays
- Confrontation
- Family therapy initially focusing on the impact of client behavior on family members

- Overt practice
- Predicting the recurrence of symptoms
- Discussing painful memories
- Accessing affective responses
- Directed imagery
- Interpretation of the transference
- Interpretation of resistance

If subjective distress is quite high, an immediate goal is to reduce the anxiety level. Doing so would be particularly urgent if the distress is sufficiently high to result in a significant disruption in the ability to cope. A wide variety of psychosocial techniques are available, mostly characterized by being supportive, structured, and designed to enhance relaxation. If a client's arousal is expressed primarily through physiological signs, techniques targeted at this level are warranted and might include:

- Progressive muscle relaxation
- Hypnotically assisted physiological relaxation
- Guided imagery
- Biofeedback
- Aerobic exercise
- Graded exposure

Arousal that is more socially or cognitively related might be reduced most effectively through these techniques:

- Meditation
- Reassurance
- Emotional support
- Cathartic discharge
- Supportive challenging of dysfunctional cognitions
- Time management
- Thought stopping

Pharmacotherapy might be useful but should be accompanied by learning new coping skills so that medication can be discontinued as soon as possible. The newly acquired coping skills then decrease the likelihood of relapse after the medication has been discontinued.

UNDERSTANDING PROBLEM CONTEXT

Coping Style

Theory, research, and clinical observations indicate that client coping style varies on a continuum between externalization and internalization. Externalizers cope with their problems by impulsively acting out, externalizing blame, attributing the cause of their

difficulties to bad luck or fate, and actively attempting to avoid their problems. They are not psychologically minded and, as a result, do not tend to respond to insight with behavioral change. In contrast, internalizers are more prone to blame themselves, based in part on the perception that they do not have the sufficient skills or abilities to overcome their difficulties. Accordingly, they tend to experience more subjective distress than externalizers. To cope with this distress, they are likely to attempt to understand their difficulties in more depth.

Clinical indicators for externalization based on history and behavioral observations include (Gaw & Beutler, 1995):

- Blaming others for their problems
- Paranoia
- Low frustration tolerance
- Extroversion
- Unsocialized aggression
- Manipulation of others
- Distraction through seeking stimulation
- Somatization with a focus on seeking secondary gains

In contrast, internalizers are more likely to have these characteristics:

- Introversion
- Intellectualization
- Constricted or overcontrolled emotions
- Minimizing difficulties
- Social withdrawal
- Somatization with symptoms related to the autonomic nervous system

MMPI-2 assessment of externalization for clinical populations can be made by finding the sum of *T* scores on 4 (Psychopathic Deviance), 6 (Paranoia), and 9 (Mania) and then comparing this with the sum of *T* scores on the internalization measures of 2 (Depression), 7 (Psychasthenia), and 0 (Social Introversion). If the sum of externalization (4 + 6 + 9) is greater than internalization (2 + 7 + 0), the client can be considered an externalizer. Conversely, if the internalizing sum (2 + 7 + 0) is greater than the sum for externalization (4 + 6 + 9), the client is likely to internalize conflicts and stress (Beutler et al., 1991, 2003; Harwood et al., 2011). Note that the preceding ratio has been designed for use with clinical populations who have at least some elevations on the MMPI-2 scales. For depressed patients, greater sensitivity can be achieved by calculating the sum of *T* scores for scales 4 (Psychopathic Deviance) and 6 (Paranoia), which should be above 125 to fulfill the criteria for having an externalizing coping style. Although not written about as much, a similar technique can be applied to the comparable scales on the MMPI-2-RF.

Several additional measures might also provide useful information related to coping style. For example, the MCMI-IV scales of Histrionic, Antisocial, Turbulent,

Aggressive/Sadistic, and Paranoid conceptually suggest externalizing styles. In contrast, the MCMI-IV scales of Avoidant, Depressive, Dependent, and Compulsive seem consistent with more internalizing styles of coping.

Clients using externalizing coping strategies have better treatment outcomes when behavioral, symptom-oriented interventions or specific techniques for building skills are used. In contrast, they do relatively poorly with techniques that attempt to enhance awareness and create insight (Beutler et al., 1991; Beutler & Clarkin, 1990; Beutler, Harwood, et al., 2006; Beutler, Harwood, Kimpara, Verdirame, & Blau, 2011; Castonguay & Beutler, 2006; Kadden, Cooney, Getter, & Litt, 1990). Techniques that are likely to be effective with externalizers include those listed next.

- Social skills enhancement
- Assertiveness training
- Group interventions
- Anger management
- Graded exposure
- Reinforcement
- Contingency contracting
- Behavioral contracting
- Questioning dysfunctional beliefs
- Practicing alternate thinking
- Stimulus control
- Thought stopping
- Counterconditioning
- Relaxation

High internalizers benefit the most from techniques that emphasize the development of insight and of emotional awareness (Beutler et al., 1991; Beutler & Clarkin, 1990; Beutler, Harwood et al., 2002; Beutler, Harwood, Kimpara et al., 2011; Castonguay & Beutler, 2006; Kadden et al., 1990). Specific techniques might include:

- insight-oriented interventions,
- therapist-directed imagery,
- direct instruction,
- outside reading (bibliotherapy),
- interpreting transference reactions,
- interpreting resistance,
- meditation, and/or
- two-chair work.

Social Support

Level of environmental support refers to the presence of a strong, cohesive family or friend network and a secure form of employment. These external means of support can

often modify the impact of other forms of stressors. High social support has also been associated with a favorable response to treatment (Mallinckrodt, 1996; Warren, Stein, & Grella, 2007), as well as the ability to maintain the gains made through treatment (Zlotnick, Shea, Pilkonis, Elkin, & Ryan, 1996). Not only are the treatment gains higher for persons with high social support; they also achieved these gains in a shorter period (Moos, 1990). In contrast, clients with low social support require more time to benefit from therapy. In general, individuals with high social support have better prognoses than those with low social support (Moos, 1990; Panzarella, Alloy, & Whitehouse, 2006).

Informal assessment of social support can be achieved by noting these characteristics:

- The extent to which the client feels trusted and respected by the people in his or her life
- The extent and quality of people he or she can confide in
- Level of experienced loneliness
- The extent the client feels abandoned by family or friends
- The extent to which the client feels a part of his or her family or friend network
- The number of friends the client has common interests with

It should be stressed that assessing social support should not consider merely the number of people available for the person, but also the quality of this support that the client feels. It is one thing to be living with a large number of people and quite another to actually feel that it is possible to confide in those people.

There are also a number of more formal strategies for assessing social support. Probably the most frequently used scale is the Social Support Questionnaire (Sarason, Levine, Basham, & Sarason, 1983). Additional information related to social support might be the relative elevation of MMPI-2 scale 0 (Social Introversion). High scores suggest an inhibited, shy person who may find it difficult to have a large network of friends. In addition, MMPI-2/MMPI-2-RF elevations on 6 (Paranoia)/RC6 (Ideas of Persecution) and 8 (Schizophrenia)/RC8 (Aberrant Experiences) suggest that both the number and the quality of social support may be low. High scores on 1 (Hypochondriasis)/RC1 (Somatic Complaints) and 3 (Hysteria)/RC3 (Cynicism) may indicate that, although the number of supports may be high, the quality of these supports may be poor. MCMI-IV elevations may also provide useful information related to social support. High scores on Schizoid, Avoidant, Schizotypal, Paranoid, and Thought Disorder each might indicate both a low number and a low quality of social support. Other scale elevations, including Dependent, Histrionic, Narcissistic, Passive-Aggressive (Negativistic), Self-Defeating, and Borderline, may have moderate to high social supports, but these supports are also likely to be quite conflicted. For example, those high on Dependent may have social supports but have achieved these supports through sacrificing their autonomy and sense of personal competence. They are also likely to experience some anxiety related to fears that this social support may not be permanent. Finally, high scores on the PAI NON (Nonsupport) scale are a clear indicator of a lack of social support. When evaluating social support, however,

cultural factors, such as whether the person comes from a highly collectivist culture, may need to be taken into account.

Low support suggests that cognitive behavioral therapy will be more effective than therapies designed to enhance relationships (Beutler et al., 2000). Both longer duration of therapy and the possibility of medication may be indicated. A supportive group intervention might be useful in providing sufficient support to activate more relationship/interpersonal types of therapies.

High social support suggests a shorter duration of therapy. Long-term intervention may even be contraindicated. Therapeutic gains are likely to occur relatively rapidly and be maintained. Therapies that enhance the quality of relationships are likely to be particularly effective, presumably because they are enhancing skills the person already has. In contrast, cognitive and behavioral therapies are likely to be less effective (Beutler et al., 2000).

Current Life Circumstances

A major consideration in what is likely causing or maintaining problems for a client is what is currently happening, personally, interpersonally, and environmentally, within his or her life. For example, an individual may present with distress that looks similar (symptomatically) to someone else, but it may be related to a recent loss (bereavement), going through a divorce, or having had a recent medical diagnosis (adjustment), or it may not have a clear environmental stimulus (perhaps a mood or anxiety disorder). Stressors in life accumulate, with more severe life stressors generally being related to more pathological responses (Parry & Shapiro, 1986). These environmental and situational stressors may necessitate the adaptation of even evidence-based treatments to meet the needs of the particular client. For example, an impending relocation may necessitate the inclusion of shorter-term or more intensive treatment techniques. Recent diagnosis of a terminal illness may require the inclusion of more palliative techniques, rather than strategies for behavior change (see Haley, Larson, Kasl-Godley, Neimeyer, & Kwilosz, 2003; Kleespies, 2004). Although many of the client characteristics discussed in this chapter are extremely important to consider, it is also important to remember that individuals do not exist outside of their multiple environmental contexts (Bronfenbrenner, 1979), each of which can contribute and moderate the problems with which individuals are struggling.

TREATMENT-SPECIFIC CLIENT CHARACTERISTICS

Resistance

Clients vary on the extent to which they are open, accepting, and responsive to treatment versus being resistant and oppositional. This resistance is frequently a defense against what they perceive as others attempting to exert or intrude on their sense of control. Those who are most resistant are likely to have a constellation of traits, including need for control, hostility, impulsivity, and direct avoidance (Dowd & Wallbrown, 1993; Shen & Dillard, 2007). They may also have difficulty accepting feedback and

may lack empathy. In addition to the preceding trait perspective, resistance can also be a state; when it is a state, rather than a trait within the person, it is known as *reactance* (Brehm & Brehm, 1981). The defensive or reactant state usually occurs when the client feels as if his or her freedom is somehow being threatened. Persons who are prone to being resistant are more likely to feel that they have a continual lack of personal control. As a result, they may compensate for this and establish a sense of control by acting in ways that oppose what is being requested or demanded of them. This is most likely to occur when the threatened area of freedom is important to the person and the individual making the request is doing so in an authoritative fashion, such as through instruction, confrontation, directives, or structured techniques. Such a structured, directive approach can potentially result in actual increases in client dysfunction. Understandably, highly reactant clients are likely to have a poorer prognosis than those who are more responsive and receptive.

Clinical indicators that may suggest high resistance include (Gaw & Beutler, 1995) those listed next.

- Extreme need to maintain autonomy
- Opposition to external influences
- Dominance
- Anxious, oppositional style
- History of interpersonal conflict
- Poor response to previous treatment
- Refusal to accept therapist interventions
- Failure to complete homework assignments

In contrast, a low level of resistance is suggested by the next indicators.

- Seeking of direction
- Submissiveness to authority
- Openness to experience
- Acceptance of therapist interventions
- Successful completion of homework assignments
- Tolerance of events beyond his or her control

Although the MMPI-2/MMPI-2-RF and MCMI-IV do not have pure measures of resistance, elevations on some scales might be consistent with high resistance. Specifically, high scores on MMPI-2/MMPI-2-RF *L/L-r* and *K/K-r* are likely to have oppositional styles, as would elevations on 6 (Paranoia)/RC6 (Ideas of Persecution) and possibly 1 (Hypochondriasis)/RC1 (Somatic Complaints). Additionally, the MMPI-2 TRT (Treatment Readiness Scale) was designed to assess treatment-specific resistance (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Beutler et al. (1991) have used a combination of the MMPI research scales for anxiety (Taylor Manifest Anxiety Scale) and social desirability (Edwards Social Desirability Scale) as a measure of resistance. MCMI-IV elevations on scales for Narcissistic, Negativistic

(Passive/Aggressive), Paranoid, Aggressive/Sadistic, and Compulsive also suggest a defensive, oppositional person. In contrast, elevations on Dependent and Histrionic suggest a more responsive, compliant style. The PAI RXR (Treatment Rejection) scale measures aspects of the person that are related to likely treatment compliance, such as openness to change and willingness to accept responsibility for problems. Beside the PAI RXR, the most frequently used pure measure of resistance (reactance) is Dowd, Milne, and Wise's (1991) Therapeutic Reactance Scale; scores above 68 indicate sufficient resistance/reactance to have implications for treatment planning.

Clients who are responsive and compliant are likely to achieve the most gains when therapists use a more directive, structured approach (Beutler, Harwood, Michelson, Song, & Holman, 2011; Beutler, Moleiro, & Talebi, 2002; Beutler et al., 1991, 1996; Castonguay & Beutler, 2006; Gaw & Beutler, 1995; Harwood et al., 2011; Horvath & Goheen, 1990). Specific techniques might include:

- Behavior contracting,
- Contingency management,
- Graded exposure,
- Direct hypnotic suggestion,
- Stimulus control,
- Cognitive restructuring,
- Developing alternative client self-statements,
- Directed imagery,
- Advice,
- Thought stopping, and/or
- Therapist interpretation.

For highly resistant clients, in contrast, strong empirical relationships have been found between positive treatment outcome and the use of nondirective, supportive, self-directed interventions (Beutler & Clarkin, 1990; Beutler, Harwood, Michelson et al., 2011; Beutler, Moleiro, & Talebi, 2002; Beutler, Engle et al., 1991; Beutler, Sandowicz et al., 1996; Castonguay & Beutler, 2006; Harwood et al., 2011). Specific techniques might include:

- Self-monitoring,
- Therapist reflection,
- Support and reassurance, and/or
- Supportive interpretation of transference.

In addition, paradoxical techniques (e.g., double binds) have been found particularly effective with reactant clients. Such techniques might include:

- Encouraging relapse,
- Prescribing that no change occur, and/or
- Exaggeration of the symptom.

Paradoxical techniques will likely be more effective if resistance levels are quite high, as might be reflected in scores above 84 (top 25%) on the Therapeutic Reactance Scale (Beutler et al., 1996; Debord, 1989; Dowd & Wallbrown, 1993; Horvath & Goheen, 1990). It is important to note that paradoxical techniques should only be employed by therapists who are specifically trained and competent to employ them.

Stage of Change

Clients undergo a series of steps during the process of change. Accordingly, any client referred for evaluation may be at a different stage in the change process. Some individuals might be simply considering the possibility of change but have not yet struggled with the specifics of how to accomplish it. This might be particularly true for involuntary referrals who are resistant and experiencing a low level of subjective distress. On the other extreme might be a client who has already taken a number of clear steps for change but is seeking help to prevent relapse. According to the stage of change, a client might require somewhat different approaches. However, considering stage of change may not be relevant for disability, medical, or many court assessments (e.g., personal injury), as facilitating change may not be part of the referral question. In these cases, assessment of the current level of functioning or differential diagnosis becomes the focus of the report.

The stages of change are likely to be quite variable. One person might pass through the different stages quite rapidly, and another who is perhaps more ambivalent or less directed might have been considering the possibility of change for years. During the process of successful therapy, it would be expected that the client would have undergone all the different stages at some point. As a result, practitioners need to be continually aware of possible alterations in the stage of change and adapt their interventions accordingly. In addition, a client might have several problem areas, especially if the problem is complex, and each area might be at a different stage in the change process. This variability requires a flexible approach depending on which area is being addressed.

Prochaska and DiClemente (1984, 2005) described five stages in the change process: (1) precontemplation, (2) contemplation, (3) preparation, (4) action, and (5) maintenance. Each stage has a different set of tasks that must be accomplished before proceeding to the next stage. The first three stages are processes that occur before any actual change or actual attempts at concrete change. In the *precontemplation* stage, people have little intention of changing behavior or attitudes. They might be vaguely aware that change needs to occur, but, for the most part, they are unaware of the possible importance of change. In contrast, other people they relate with often clearly see the need for change. As a result, these clients are likely to be referred or seek treatment when the legal-justice system threatens to punish them, a spouse threatens to leave them, parents threaten to disown them, or an employer threatens to dismiss them. Under these conditions, change is likely to proceed only if there is either continual outside pressure or the client internalizes the need for change. When individuals begin to consider change more seriously, they can be considered to be in the *contemplation* stage. At this point, they are aware that they have a problem and are concerned with

how coping with the problem might best be accomplished. However, they have not yet committed themselves to the process. In the *preparation* stage, individuals have become more committed to change, which is represented by their intent to take action in the near future. This intent may also be accompanied by the possible presence of minor experiments with new behaviors. Because they are not yet clear on how best to accomplish their intended change, these individuals may need help considering all relevant options and choosing the optimal strategy for implementing the change.

The final two steps in the change process focus on actually implementing the change and ensuring that it is maintained. *Action* is the point at which clients actually change their environment, thoughts, attitudes, or behavior. Doing this often requires a considerable amount of time and energy, and, as a result, individuals must be highly committed. At this point, changes are most clearly visible to others. The preceding preparatory and contemplative processes should not be underestimated, however, because they are crucial in determining the relative success of any change. During the *maintenance* stage, individuals work to consolidate change and prevent relapse.

These interview questions can help determine the individual's stage of change: Do you intend to change in the near future? Are there current changes you are going through? Have you made changes? Are you currently working to prevent relapse? These questions might be incorporated into an intake form or interview (Prochaska, Norcross, & DiClemente, 1994). It may be necessary to probe or otherwise obtain clarification to clearly determine the stages of change. Formal assessment of the stages of change can be made using a few different measures. These include the 32-item Stages of Change Scale, also called the University of Rhode Island Change Assessment (URICA; McConaughy, DiClemente, Prochaska, & Velicer, 1989; McConaughy, Prochaska, & Velicer, 1983), which assesses stage of change for general problems, and the 19-item Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES; Miller & Tonigan, 1996), which focuses on stage of change related to stopping drug and alcohol use.

Research has generally supported the clinical utility and predictive validity of tailoring interventions according to the different stages of change. This research has focused primarily on problems such as addictive behaviors, weight control, eating disorders, sunscreen use, and initiating an exercise program (Geller, Cockell, & Drab, 2001; Norcross, Krebs, & Prochaska, 2011; Prochaska, 2000; Prochaska, DiClemente, & Norcross, 1992; Prochaska, Rossi, & Wilcox, 1991). While research in other areas is promising, more extensive research needs to be conducted to determine the effectiveness of tailoring interventions to stage of change with a wider range of problem areas (Norcross et al., 2011; Whitelaw, Baldwin, Bunton, & Flynn, 2000). In the areas researched, there is generally evidence that tailoring interventions toward the stage of change can optimize treatment outcome (Norcross et al., 2011; Petrocelli, 2002; Prochaska, 2000; Prochaska & DiClemente, 2005; Prochaska et al., 1992). Following is a summary of the stages of change and associated intervention recommendations:

- *Precontemplation Stage.* This stage is often, although not necessarily, consistent with involuntary referrals. As a result, resistance level may be high and subjective distress low, such that interventions would need to be made accordingly (e.g., increase arousal; use nondirective, supportive techniques and paradoxical

interventions). Because these clients might feel ambivalent about treatment, it is crucial to spend time building rapport and discussing areas that work or do not work in their lives. In general, interventions that increase clients' awareness and understanding of the potential benefits of changing are most effective for helping them move from precontemplation to contemplation.

- *Contemplation and Preparation Stages.* As in the previous stage, enhancing the relationship with the client is particularly important. Providing understanding and awareness is also crucial. Doing this should include exploring the client's interpersonal or behavioral patterns, reasons for and against changing, and the different strategies for creating change. An inventory of client strengths or resources and weaknesses might also be useful. The first three stages might be most consistent with humanistic or psychodynamic approaches that stress insight, exploration, value clarification, clarification of personal goals, novel experiences, and cathartic experience. Specific strategies in these stages include evaluating the client's feelings about changing, including feelings about not changing, and ultimately strategies that help the client commit to and feel confident to act.
- *Action Stage.* A wide variety of specific, concrete techniques might be used in this stage. The selection of techniques depends in part on the problem itself, functional impairment, problem complexity, subjective distress, and resistance. Specific strategies can be implemented that might involve changes in concrete behavior, patterns of interpersonal relationships, self-statements, or ways of experiencing the world. Cognitive or behavioral techniques might be most effective at this point, particularly stimulus control, graded exposure, cognitive restructuring, role plays, social skills training, or counterconditioning. Restructuring the client's life to avoid stimuli that elicit the problem behaviors, as well as practicing alternative behaviors for the problem behaviors can be bolstered by efforts to reinforce these positive changes.
- *Maintenance Stage.* At this point, the therapist can become like a coach or a consultant who advises and encourages the client. A crucial consideration is how relapse is most likely to occur and to develop countermeasures to prevent these situations from occurring or at least to minimize their impact over a longer period. Specific techniques might include stimulus control, social contracting, enhancing social support, anger management, or a behavioral contract requiring the person to take preventive measures if relapse seems likely.

Other Client Characteristics

Other client characteristics have been identified as important for the clinician making specific decisions about recommended treatment options. Among these are the culture of the individual being assessed, as well as his or her expectations of treatment and preferences in treatment options. It can be overwhelming to try to match treatment recommendations to every single client characteristic, but some characteristics may be more salient in some circumstances than in others. For example, a client who speaks a language other than English as a primary language would likely best be served in

his or her primary language. However, language and culture may not be as salient and important for an English-speaking, heterosexual, middle-class white man. Which client characteristics to pay close attention to will depend on the specifics of the client being assessed and on the findings of the assessment.

One major client characteristic to consider when making treatment recommendations is the client's cultural background. Language, ethnicity, religion, economic background, sexual orientation, and other factors contribute to the complexity of culture for clients. As such, it would be overwhelming, inefficient, and impractical to identify all the factors that should be included in any one grouping of these variables. Instead, the literature has focused on adapting therapies based on cultural characteristics (e.g., Barrera & González Castro, 2006; Hwang, 2009; Whitbeck, 2006). Cultural adaptations of treatments have been found to be especially impactful on therapeutic outcome for Asian American clients, with lesser but still important effects with African American and Latino/a clients, and even lesser but still impactful effects with Native American clients (T. B. Smith et al., 2011). These culturally adapted treatment effects are especially important for adult and older adult clients, as opposed to child, adolescent, and young adult clients. Much of the literature has focused on ethnic minorities, but models have also been proposed more recently for adapting treatments for sexual and gender minorities (Austin & Craig, 2015), religious clients (though with less consistent positive effect; Lim, Sim, Renjan, Sam, & Quah, 2014; Worthington, Hook, Davis, & McDaniel, 2011), and low socioeconomic status (Miranda, Azocar, Organista, Dwyer, & Areane, 2003), as well as others. Cultural adaptation of treatments may take different forms; some have proposed flexibility in the approach to evidence-based treatment, in order to alter interventions as needed based on what emerges from the client (Kendall & Beidas, 2007), whereas others have proposed more systematic, specific alterations of manualized treatments (e.g., G. C. N. Hall, 2001; Sue, Bingham, Porché-Burke, & Vásquez, 1999; Trimble & Mohatt, 2006). Culture may be one consideration when making specific recommendations from assessment results.

More and more, researchers are evaluating the benefit of taking into consideration client preferences when recommending treatment, a move toward increasing inclusion and collaboration with clients in a decision that will likely have great impact on them. These preferences have been categorized into three broad domains: (1) role preferences, which include the role clients would like to be in and the role they would like their therapist to play; (2) client preferences about the therapist, including traits like therapist ethnicity, gender, years of experience, or personality attributes; and (3) the client's treatment preferences, which include theoretical orientation or modality of treatment (Swift, Callahan, & Vollmer, 2011). Swift, Callahan, and Vollmer's (2011) meta-analysis revealed that matching treatment to clients' preferences along these three dimensions decreased drop-out from treatment by a third and had a medium effect size on treatment outcome. The positive effect of matching preferences for treatment was found mostly for those with anxiety, depressive, and substance abuse disorders, rather than for those with health problems or serious mental illness.

In addition to treatment preferences, the idea of client expectation for the effectiveness of psychotherapy has become important. The positive expectation for treatment has been found to have a small but significant impact on the outcome of treatment (Constantino, Glass, Arnkoff, Ametrano, & Smith, 2011). This finding is

more important for the delivery of feedback from assessment than it is in making clinical decisions for recommendations. That is, if the feedback and recommendations presented to the client are convincing, such that they increase the client's confidence about the likely effectiveness of the recommended treatment, it may contribute to more success in the actual outcomes.

THE SYSTEMATIC TREATMENT SELECTION (STS)/INNERLIFE APPROACH

The preceding dimensions are intended to be logically consistent as well as manageable. However, it is difficult, especially for less experienced practitioners, to internalize and use every single one of these dimensions when making clinical decisions on the basis of a psychological assessment. With practice, it is likely that many of the features will become progressively more internalized, perhaps even requiring less formal assessment. Clinicians may prioritize some of the dimensions above others, depending on the specific details of the client with whom they are working. Additionally, as further research provides more precise definitions of empirical associations, additional dimensions will likely be included. Each of these developments will bring clinicians closer to Paul's (1967) ultimate goal of combining the best treatment with the optimal mix of therapist, client, problem, and context.

In order to systematize and simplify the consideration of many of the dimensions already discussed, Beutler and his colleagues (Beutler, 2011; Beutler & Clarkin, 1990; Beutler et al., 2000; Beutler et al., 2006; Harwood & Williams, 2003) developed a model of treatment selection based primarily on the identification of many of these client characteristics, called Systematic Treatment Selection (STS). This approach relies on systematically identifying these characteristics and making recommendations based on empirically and clinically established associations with treatment outcomes. These characteristics include degree of functional impairment, social support, level of problem complexity/chronicity, and subjective distress/motivation, with special emphasis on coping style and resistance (Beutler, 2011). This model not only closely adheres to evidence-based research and can include many of the assessment techniques discussed in previous chapters; it also follows a clear sequence of decision making and is comprehensive, while detailing a manageable number of variables.

When a practitioner is confronted with a client, he or she should acquire the relevant information and, based on this information, should develop a series of decisions and recommendations. Beutler and his colleagues (Beutler et al., 2000; Beutler & Harwood, 2000; Beutler et al., 2006; Harwood et al., 2011; Harwood & Williams, 2003) have incorporated at least seven client dimensions and related these to different types of decisions (see Table 14.1). The first variable relates to *functional impairment* and has clear implications for general case management.

As noted in the earlier section titled "Functional Impairment," some of these issues include the relative restrictiveness of therapy (inpatient/outpatient), whether medication should be considered, the intensity of treatment (duration and frequency), and the immediate goals of therapy. The other six dimensions relate more to specific techniques of intervention than to general case management.

Table 14.1 STS Characteristics Included in Treatment Planning

Variable	Treatment considerations
1. Functional impairment	Restrictiveness (inpatient/outpatient) Intensity (duration and frequency) Medical versus psychosocial interventions Prognosis Urgency of achieving goals
2. Social support	Cognitive behavioral versus relationship enhancement Duration of treatment Psychosocial interventions versus medication Possible group interventions
3. Problem complexity/chronicity	Narrow symptom focus versus resolution of thematic unresolved conflicts
4. Coping style	Behavioral symptom-oriented versus internal insight-oriented interventions
5. Resistance	Supportive, nondirective, or paradoxical versus structured, directive interventions
6. Subjective distress	Increase/decrease arousal
7. Stage of change	Differential short-term treatment goals/targets Supportive versus insight-orientation versus cognitive/behavioral interventions

Level of *social support* can be used to determine whether a client's social network can be relied on or whether it needs to be increased. The relative *complexity* (and *chronicity*) of a client's problem is important in considering whether the focus of treatment should be on specific, discrete, environmentally related symptoms or more on internal, chronic areas of conflict. In addition, *coping style* can help guide whether interventions should be on changing external behavior or directed at more internal, insight-oriented levels of change, and level of *resistance* (reactance) has implications for how directive interventions should be. *Subjective distress* can be used to guide clinicians in determining whether the client's level of arousal should be increased or decreased. The *stage of change* the client is in also helps determine categories of specific techniques that are likely to be useful.

Each of these dimensions can potentially be assessed with a combination of formal tests, interview data, behavioral observations, and relevant history, ranging from a relatively short interview that focuses on each of the relevant domains to an extensive battery consisting of a number of formal psychological tests. In addition, a rating scale (the *STS Clinician's Rating Form*; Corbella et al., 2003; Fisher, Beutler, & Williams, 1999) has been developed to assist in summarizing the various ratings. The *Psychological Report Writing Assistant* (Groth-Marnat & Davis, 2014) provides a framework for recommendations based on the STS. Completing the program takes between 20 and 40 minutes. Finally, clients (and clinicians) can enter data online at www.innerlife.com (Beutler, Williams, & Norcross, 2008) so that the STS dimensions can be rated based on

client responses and a treatment plan can be developed. The STS offers a thorough, systematic way to make decisions related to treatment and other recommendations using multiple client variables.

RECOMMENDED READING

- Antony, M. M., & Barlow, D. H. (2011). *Handbook of assessment and treatment planning for psychological disorders* (2nd ed.). New York, NY: Guilford Press.
- Beutler, L. E., & Clarkin, J. F. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions*. New York, NY: Guilford Press.
- Butcher, J. N., & Perry, J. N. (2008). *Personality assessment in treatment planning: Use of the MMPI-2 and BTPI*. New York, NY: Oxford University Press.
- Groth-Marnat, G., Roberts, R., & Beutler, L. E. (2001). Client characteristics and psychotherapy: Perspectives, support, interactions, and implications. *Australian Psychologist*, 36, 115–121.
- Harwood, T. M., Beutler, L. E., & Groth-Marnat, G. (2011). *Integrative assessment of adult personality* (3rd ed.). New York, NY: Guilford Press.
- Hersen, M., & Sturmey, P. (2012). *Handbook of evidence-based practice in clinical psychology* (Vols. 1, 2). Hoboken, NJ: Wiley.
- Jongsma, A. E., Peterson, L. M., & Bruce, T. J. (2014). *The complete adult psychotherapy treatment planner: Includes DSM-5 updates* (5th ed.). Hoboken, NJ: Wiley.
- Norcross, J. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed.). New York, NY: Oxford University Press.
- Norcross, J., Santrock, J. W., Campbell, L. F., Smith, T. P., Sommer, R., & Zuckerman, E. L. (2012). *Authoritative guide to self-help resources in mental health* (4th ed.). New York, NY: Guilford Press.
- Prochaska, J. O., & DiClemente, C. C. (2005). The transtheoretical approach. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed.; pp. 147–171). New York, NY: Oxford University Press.
- Wright, A. J. (2010). *Conducting psychological assessment: A guide for practitioners*. Hoboken, NJ: Wiley.

THE PSYCHOLOGICAL REPORT

The psychological report is one of the most important end products of assessment. It represents the clinician's efforts to integrate the assessment data into a functional whole so that the information can help the client improve his or her life, helping to solve problems and make decisions. Even the best tests are useless unless the data from them are explained in a manner that is relevant and clear and that meets the needs of the client and referral source. Doing this requires clinicians to give not merely test results, but also to interact with their data in a way that makes the conclusions useful in answering the referral questions and making recommendations.

An evaluation can be written in many different ways. The manner of presentation depends on the purpose for which the report is intended as well as on the individual style and orientation of the practitioner. The format provided in this chapter is merely a suggested outline that follows common and traditional guidelines. It includes methods for elaborating on essential areas, such as the referral question, behavioral observations, relevant history, impressions (interpretations), and recommendations. This format is especially appropriate for clinical evaluations that are problem-oriented and that offer specific prescriptions for change. Additional alternatives for organizing the report are to use a letter format, give only the summary and recommendations, focus on a specific problem, summarize the results test by test (though this is discouraged, as discussed later in this chapter), write directly to the client, or provide client descriptions around a particular theory of personality. The sample evaluations vary somewhat from the suggested format, although they usually still include the essential categories of information that are discussed in this chapter.

An important note is that the quality and usefulness of a report is predicated on the assessment itself being thorough, valid (accurate), and clear. A well-written report cannot compensate for poor testing. In addition to ethical, reliable, and valid processes for conducting an assessment, reports and recommendations are improved by the practitioner being knowledgeable about the area or type of issue the client is experiencing. Such knowledge helps to increase the depth of the interpretations and provides relevant information or a general map of the problem area that can be used to help ensure that all relevant aspects have been covered. Importantly, background knowledge about the problem area provides relevant information on a range of interventions, as well as on the effectiveness of these interventions. For example, knowledge regarding depression means that the practitioner is aware of its causes, the variety of ways in which it is expressed, options for interventions, and when further assessment is indicated (e.g., for suicide potential). Often consulting a well-written, up-to-date chapter will provide sufficient information. In the general clinical area, useful resources are the *Clinical Handbook of Psychological Disorders* (Barlow, 2014), *Clinician's Handbook*

of *Evidence-Based Practice Guidelines* (O'Donahue & Fisher, 2006), or *Pocket Handbook to Clinical Psychiatry* (Sadock & Sadock, 2010). Persons doing cognitive evaluations might consult *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration* (Groth-Marnat, 2000a), *Neuropsychological Assessment* (Lezak, Howieson, Bigler, & Tranel, 2012), or *Clinical Neuropsychology: A Pocket Handbook for Assessment* (Parsons & Hammeke, 2014). Educational report writers might benefit from reading relevant sections in *Assessment of Children* (Sattler, 2008; Sattler & Hoge, 2006); *Foundations of Behavioral, Social, and Clinical Assessment of Children* (Sattler, 2014); or *Handbook of Evidence-Based Practice in Clinical Psychology, Child and Adolescent Disorders* (Hersen & Sturmey, 2012). A useful resource when doing vocational assessments is *The Essentials of Career Interest Assessment* (Prince & Heiser, 2000).

GENERAL GUIDELINES

In 2010, the American Psychological Association recognized Personality Assessment as a specific and distinct proficiency, which means it is a specific, defined skill set within the practice of psychology that can be evaluated, measured, and ultimately verified in order to protect the public from poor quality or even harmful services. As part of the proficiency process, practitioners can apply to be recognized for their personality assessment proficiency, and the Society for Personality Assessment (SPA) took the leadership in determining the process for recognition. The core of the recognition process includes the evaluation of a written psychological assessment report (as a proxy for the overall assessment process). The report review form and process (rubric) developed by SPA applies to personality assessment, but the concepts apply to all of assessment. This section of the chapter is organized around the five major themes presented in the proficiency rubric: comprehensiveness, integration, validity, client-centeredness, and overall writing. The rubric, which can be used to guide training and improve assessment reports, is presented in Figure 15.1.

Comprehensiveness

Length

The typical psychological report is between 5 and 7 single-spaced pages (Groth-Marnat & Horvath, 2006). However, the length can vary substantially based on the purpose of the report, context, and expectations of the referral source. In medical contexts, a 2-page report is not uncommon. This parallels the format of many physician reports that are a similar length. In contrast, legal contexts often require reports that are from 7 to 10 pages because of the greater need for documentation combined with more extensive referral questions. It is not unusual for a psychologist serving as an expert witness to not only evaluate a client, but also anticipate and defend him- or herself against rebuttals, as well as comment on reports made by other mental health professionals. The more moderate (and frequent) 5- to 7-page report is particularly prevalent in psychological, educational, and vocational contexts. Although this length is fairly typical in clinical practice, one of the most frequent complaints related to psychological

PERSONALITY ASSESSMENT PROFICIENCY: REPORT REVIEW FORM

Applicant Name: _____

Reviewer Name: _____

Date: _____

- I. Please consider each criteria item as either:
Met proficiency criterion (Yes, circle 1 point) or
Not met proficiency criterion (No, circle 0 points).
- II. Critical items are noted with an asterisk (*).
- III. Please include any comments you may have regarding each section (positive & constructive feedback) and overall proficiency.

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
COMPREHENSIVENESS			
1. Adequate and appropriate identifying information is presented.	The report includes basic demographic information about the individual and relevant current circumstances.	0	1
2. The referral source is clearly identified.	The report specifically indicates the origin of the referral.	0	1
3. A referral question or reason for the assessment is clearly stated.	The report clearly states the reason for the evaluation, so that it can be determined if the purpose, conclusions, and recommendations are aligned and the referral question is ultimately addressed.	0	1
4. The history provided is adequate/relevant to the assessment question(s).	The history presented allows the reader to contextualize the referral questions, presenting problems, conclusions, and recommendations within the context of the individual and his/her culture. If the reviewer feels there is too much (or additional, irrelevant) history included in the report, the report should still be considered to meet this criterion, unless there is <i>substantial</i> irrelevant data, which detracts from the clarity of the report. Special attention should be paid to salient omissions or missing information that would be important in determining the validity and applicability of the test results to the individual situation of the person being evaluated.	0	1

Figure 15.1 Personality Assessment Proficiency Review Report Form

Source: Reproduced by special permission of the Society for Personality Assessment (SPA). Further reproduction is prohibited without permission of SPA.

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
5. Observations of the client's behavior and engagement in the assessment are presented.	The report includes a discussion, however brief, of the likely engagement and effort in the process, often based on behavioral observations. Behavioral observations can also inform other areas of the assessment, serving as supportive or contradictory evidence together with test data.	0	1
6. A summary section is included.	The report provides a clear (and usually succinct) summary of impressions that integrates the history, behavioral observations, and test results to address the referral question.	0	1
<i>Comments on Comprehensiveness:</i>			
INTEGRATION			
*7. The assessment includes at least three different assessment methods (e.g., self-report, performance-based, clinical interview).	The report utilizes a minimum of three types of assessment measures/resources toward assessing an individual's personality/emotional functioning. While additional measures of cognitive or other specific areas of functioning may be included, this criterion relates to the use of measures toward understanding and explaining personality functioning.	0	1
8. Cross-method interpretations are presented in an integrated manner.	The report presents findings in a way that does not ultimately require the reader of the report to "do the work" of integrating results from disparate methods of evaluation. That is, similar (or contradictory) findings from different methods are, somewhere in the report, integrated in a way that explains the ultimate conclusions.	0	1

Figure 15.1 (Continued)

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
9. Conflicting findings are adequately addressed (if applicable).	<p>The report presents conflicting findings in a way that helps the reader understand how he/she should interpret the evidence.</p> <p>Examples include (but are not limited to):</p> <ul style="list-style-type: none"> • Explaining why different methods may yield different information (“While the client self-reported X, when evaluated using a measure that does not rely on self-report, it was revealed that Y. It is likely that this difference is due to Z.”) • Using some other result to help determine which pieces of the contradictory results should be given more credence (“Because it was found that he tends to try to paint himself in a very positive light, ...”) • Explaining the nuanced differences that mean that the seemingly contradictory information is not in fact contradictory (“Although it seems that X and Y are contradictory, in fact it is possible for someone to be both X and Y, as these traits ...”) <p><i>If there are no instances of conflicting findings, give credit for this criterion.</i></p>	0	1
Comments on Integration:			
VALIDITY			
10. The validity of test findings and quality of data are discussed.	The report acknowledges potential limitations of measures used due to diversity or other factors. Measures with weaker psychometric foundations or lacking in relevant normative data are recognized as such in some manner (direct discussion of such issues, less emphasis in discussion, etc.).	0	1

Figure 15.1 (Continued)

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
*11. Test interpretations are consistent with the empirical literature and accepted clinical practice.	The narrative descriptions of test results in the report are generally consistent with what is known in the literature and what is generally accepted clinical practice. The report presents overall what would be considered evidence-based and generally-accepted interpretations of tests. Any major variations from generally accepted practice are clearly, logically, and defensibly justified (for example, elevations on certain scales of a self-report are discussed accurately and appropriately versus over-pathologizing or not recognizing areas of concern based on testing data).	0	1
*12. Assertions made from test results are consistent with the data collected.	Using the appendix of test scores to evaluate, the report presents findings that are in fact based on the entirety of data collected. That is, no major test results are omitted for any reason, the narrative explanation of what test results mean are consistent with the actual test results/scores, and the narrative explanations of all results are not in any way misleading to a reader.	0	1
13. Test interpretations are sensitive to issues of culture and diversity, including ethnicity, race, gender, age, sexual orientation, age, religion, ability, etc.	Where appropriate, diversity issues are addressed, relating to test interpretations and overall interpretations of the evaluation. No clear and egregious instances of culturally inappropriate interpretations or assertions are presented in the report. Please note that culture reflects more than one's ethnicity and covers a wide array of diversity factors.	0	1
*14. Diagnostic impressions and conclusions are reasonable based on the data presented.	The report presents conclusions that are reasonable based on the history, test results, behavioral observations, culture, and any other relevant information. It is clear that the assessment fully justifies the conclusions.	0	1
<i>Comments on Validity:</i>			

Figure 15.1 (Continued)

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
CLIENT-CENTERED			
*15. The referral question(s) is addressed adequately.	The conclusions and recommendations are tied back to the referral question, showing that the ultimate purpose of the assessment is consistent with why it was undertaken in the first place. If the conclusions and recommendations do not address the referral question, the report should explain this clearly.	0	1
16. Overall, individual test results are presented in a way that is clearly and specifically about the individual being evaluated.	In order to remain client-focused (and to avoid making overly general statements and recommendations that could apply to most individuals), the report uses language and organization that tailors conclusions to the specific individual being evaluated. Report language is not directly copied from computer reports.	0	1
17. Overall, the report is “person-focused” rather than “test-focused.”	While many different formats are acceptable for proficiency, in general, the test report is clearly focused on the individual being assessed. Some examples include (but are not limited to): <ul style="list-style-type: none">• Presenting results from tests thematically, rather than test-by-test• Presenting results in terms of abilities or traits, rather than in terms of tests themselves• When results are presented by test or method, including a comprehensive, integrative summary that describes what the data mean, taken together, for the specific individual being evaluated	0	1
18. Recommendations flow directly and clearly from the data, including the test findings, client’s clinical presentation, referral question, and history.	The recommendations presented in the report logically stem from and are justified by the information gathered and presented during the evaluation. There is alignment between the recommendations and the conclusions drawn from the different data sources. <i>If recommendations are appropriate for the report, they may be included in answer to referral question/summary section</i>	0	1

Figure 15.1 (Continued)

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
19. Recommendations are clear, specific to the person, and reasonable.	<p>The recommendations are:</p> <ul style="list-style-type: none"> • Clear enough for the reader to be able to easily understand them • Specific to the individual being evaluated, including enough detail to increase the likelihood of success (e.g., not just recommending “therapy,” but specifying a specific type of therapy that is likely to be most helpful, and perhaps even a specific professional) • Reasonable, including being attainable by the client given his or her current circumstances (e.g., recommending a highly technical and specified treatment, which may be evidence-based, but is not reasonably attainable by a client in a rural area or with economic challenges, would not be considered reasonable). <p><i>If recommendations are appropriate for the report, they may be included in answer to referral question/summary section</i></p>	0	1
Comments on Client-Centered:			
OVERALL WRITING			
20. Test scores and response examples are appropriately used (if applicable).	<p>When and if test scores and response examples are included in the report, they serve the purpose of enhancing the reader’s understanding of the material being presented, rather than distracting from it.</p> <p><i>If there are no test scores or response examples used, give credit for this criterion.</i></p>	0	1
21. The report is clear, coherent, and generally jargon-free.	Overall, the report is written using language that is professional and minimizes jargon, such that the reader (most often including the client) will likely be able to understand and follow it easily.	0	1

Figure 15.1 (Continued)

Criterion	Annotation/Details	Met Criterion? (No = 0 Yes = 1)	
22. Overall, the report is well written, organized, and mostly free from grammatical errors.	Overall, the report is written with clear organization and free of grammatical errors.	0	1
<i>Comments on Overall Writing:</i>			
<i>Additional General Comments:</i>			



OVERALL PROFICIENCY			
Check if these items <u>met</u> the criteria:	CHECK HERE (X)		
	7		
	11		
	12		
	14		
	15		
Check if total points is 19 or MORE:			
	Are all six of these highlighted boxes checked? If so, check here:	Meets Proficiency?	Application Meets Proficiency  Application Does Not Meet Proficiency 
<p><i>An early draft of this form was developed by Mark Blais and has since been modified by the SPA Proficiency Committee into the current version. The Proficiency Report Review Form has been approved by the SPA Board of Trustees.</i></p>			

Figure 15.1 (Continued)

reports is that they are too long (Brenner, 2003). The length should depend on ensuring comprehensiveness, including all necessary and relevant information in a report that is necessary, without including too much, extraneous, irrelevant, or even distracting information.

There are no specific rules to follow in determining how much information to include in a report. A general guideline is to estimate how much information a reader can realistically be expected to assimilate. If too many details are given, the information may begin to become poorly defined and vague and, therefore, lack impact or usefulness. When clinicians are confronted with a great variety of data from which to choose, they should not attempt to include it all. A statement such as “The client’s relative strengths are in abstract reasoning, general fund of knowledge, short-term memory, attention span, and mathematical computation” is likely to overload the reader with too many details. The clinician should instead adequately develop each of the various points and focus on the areas that are most relevant to the purpose of the report.

Topics

There is an extremely wide range of topics or domains that clinicians may decide to discuss in their reports. These topics serve as conceptual tools that enable report writers to give form and direction to the information they are trying to communicate. The four most common topics are likely to be related to cognitive functioning, emotional functioning (affect/mood), self-concept, and interpersonal relations. Many reports can be adequately organized around these four domains. Additional topics include personal strengths, vocational aptitudes, suicidal potential, defenses, areas of conflict, behavior under stress, impulsiveness, or sexuality. Often an adequate report can be developed by describing just a few of these topics. For example, a highly focused report may elaborate on one or two significant areas of functioning, whereas a more general evaluation may discuss seven or eight relevant topics. Table 15.1 is a representative list of topics that may be considered for inclusion in an evaluation. This list is by no means complete but can provide a general guide to the wide range of possible topics from which a report writer can choose.

Deciding What to Include

The general purpose of a psychological evaluation is to provide information that will be most helpful in responding to the referral question and meeting the needs of the client. In this context, the clinician must strike a balance between providing too much information and providing too little and between being too cold and being too dramatic. As a rule, information should be included only if it serves to address the referral question and increase understanding of the client. For example, descriptions of a client’s appearance should be oriented toward areas such as his or her level of anxiety or resistance. A client might be described as hesitant in his or her approach to tasks and as saying something like “Why do I have to take all these tests anyway?” If the person were dressed in bizarre clothes and his or her hair were unkempt or dyed purple, this information might also be important to include if relevant to the referral question.

Table 15.1 Examples of General Topics Around Which a Case Presentation May Be Conceptualized

Achievement	Drives, dynamics	Personal consequences of
Affect	Emotional controls	behavior
Aggressiveness	Emotional functioning	Placement prospects
Antisocial tendencies	Fixations	Problem complexity
Anxieties	Flexibility	Psychopathology
Aptitudes	Frustrations	Rehabilitation needs
Attitudes	Functional impairment	Rehabilitation prospects
Aversions	Goals	Resistance
Awareness	Hostility	Sentiments
Behavioral problems	Identity	Sex
Biological risk factors	Intellectual controls	Sex identity
Cognitive functioning	Intellectual levels	Sex role
Cognitive skills	Interests	Significant others
Cognitive style	Interpersonal relations	Situational factors
Competency	Interpersonal skills	Social consequences of
Conflicts	Irrational cognitions	behavior
Content of consciousness	Lifestyle	Social role
Coping style	Mood	Social structure
Defenses	Needs	Social support
Deficits	Outlook	Special assets
Developmental factors	Perception of environment	Strengths
Diagnostic considerations	Perception of self	Subjective feeling states

Source: Adapted from *Psychological Report Writing* (3rd ed., p. 120) by N. Tallent, 1988, Englewood Cliffs, NJ: Prentice-Hall.

Generally, however, information regarding the types of clothing the person is wearing or color of his or her eyes or hair is not relevant.

The basic guidelines for deciding what to include in a report relate to the needs of the referral setting, background of the readers, purpose of testing, relative usefulness of the information, and whether the information describes unique characteristics of the person. After these general guidelines have been considered, the next step is to focus on and organize the information derived from the tests. For example, if a general review of aspects of personality is the purpose of the report, a clinician can look at each test to determine what information it can provide specifically about personality.

A further general rule is that information should focus on the client's unique method of psychological functioning. A reader is concerned not so much with how the client is similar to the average person as in what ways he or she is different. A common error in psychological reports is the inclusion of generalized statements that are so vague, they could apply to the majority of the population. These vague, generalized statements

are likely to be unconditionally accepted as applying to a person even though they are randomly selected. For example, Sundberg (1955) administered a “personality” test to a group of students and gave them all identical “interpretations” based on universal or stereotyped personality descriptions composed of 13 statements, such as:

- You have a great need for other people to like and admire you.
- You have a tendency to be critical of yourself.
- You have a great deal of unused capacity you have not turned to your advantage.
- While you have some personality weaknesses, you are generally able to compensate for them.
- At times, you have serious doubts as to whether you have made the right decision or done the right thing.

Virtually all students used in the study reported that the evaluation statements were accurate descriptions of themselves. Other studies suggest that not only were students unable to discriminate between fictitious and genuine feedback, but they may even have preferred generalized fictitious results, particularly if they were framed within a positive context (Dies, 1972). This uncritical acceptance of test interpretations might be even further encouraged when objective-appearing, computer-generated interpretations are used (Groth-Marnat & Schumaker, 1989). W. G. Klopfer (1960) referred to this uncritical acceptance of universally valid statements as the “Barnum effect,” in reference to P. T. Barnum’s saying, “There’s a fool born every minute.” The Barnum effect also speaks to the power of psychological assessments, as consumers are poised and ready to accept feedback, no matter how vague, as true. Although universal statements may add to the “subjective” validity of the report when read by the client, such statements should be avoided in favor of stressing the person’s essential uniqueness.

Summary

Because of the lengthiness of most psychological assessment reports, it is customary to include a summary toward the end of the report (often right before the recommendations). Two key considerations should be given to any summary. First, it should be comprehensive. The summary should include the most relevant, important findings from the assessment, as related to the referral questions, leaving out superfluous information. If four areas were assessed and written about in the report (e.g., cognitive functioning, emotional functioning, self-concept, and interpersonal relations), then the summary should cover each of these four areas. These four areas, whenever possible, should be integrated into a coherent narrative of the individual. The second consideration for a summary is that the writer should assume that it will be the only section read by certain people. This further emphasizes the need for it to be comprehensive, but it also means that the summary should be relatively concise. A summary that is 3 pages long will lose the interest and attention of a reader who has skipped much of the report in the first place. Ensuring that the summary is both comprehensive and concise is often a difficult task, but it may make the difference between a client truly understanding the assessment versus not.

Integration

Presenting Test Interpretations

Reports should organize information around specific referral questions. The result will be a report that is highly focused, well integrated, and avoids any extraneous material. For example, if a referral source asks whether the client is experiencing cognitive impairment, the interpretations based on the test data are directed toward answering whether this hypothesis is supported.

The interpretations should be organized around specific domains, such as coping style, memory, personality, or interpersonal relations. This approach is comprehensive, indicates the client's strengths and weaknesses, and typically gives the reader a good feel for the person as a whole. The referral question is still answered but is addressed by responding to specific domains relating to the referral question. Readers tend to prefer and better comprehend integrated reports written by addressing functional domains rather than test scores. The weakness of domain-oriented reports lies in the potential to provide too much information, thus overloading the reader.

Some reports present the results test by test, one at a time (WAIS-IV, Bender, MMPI-2, etc.). This approach clarifies the source of the data and enables the reader to understand more clearly how the clinician made his or her inferences. It is also relatively easy for the examiner to organize the results. These advantages are offset by significant disadvantages. The emphasis on tests can distract the reader and tends to reduce the client from a person to a series of test numbers. Readers of reports, regardless of their theoretical or disciplinary background, do not respond well to this style of report writing (Mendoza, 2001; Tallent 1992, 1993). A test-by-test presentation also reflects a failure to integrate the data. A test-by-test format may be a particular issue because it is not uncommon for inconsistencies to occur among different test scores. Often only half of all possible interpretations listed in an interpretive manual or computer narrative are actually true for a particular client. It is up to the clinician to determine which of these do or do not apply for the person. Sometimes a report writer using a test-by-test approach hedges his or her "interpretations" by using a phrase such as "Other persons with similar test profiles have the following qualities." The referral source, however, does not want to know about *other* people but is concerned with *this* client, at this time, living in a certain context. A test-by-test interpretation, then, suggests that the practitioner has neither adequately conceptualized relevant dynamics nor fully understood the area under investigation (Mendoza, 2001; Wolber & Carne, 1993). It also encourages the belief that an examiner is a technician who merely administers tests rather than a clinician who uses multiple sources of information to answer referral questions and helps people solve problems they are facing. This places the responsibility of reconciling conflicting findings and deciding which findings apply to the client and which do not on the reader of the report, who most often is not competent to make these interpretations. The existing literature is unanimous in discouraging a test-by-test style and, instead, strongly recommends an integrated, case-focused, problem-solving style (Groth-Marnat, 2003; Groth-Marnat & Horvath, 2006; Harwood, Beutler, & Groth-Marnat, 2011; S. Kvaal, Choca, & Groth-Marnat, 2003; Mendoza, 2001; Michaels, 2007; Sattler, 2008, 2014; Sattler & Hoge, 2006; Tallent, 1992, 1993; Wolber & Carne, 1993; Wright, 2010; Zuckerman, 2005).

Wright (2010) presented a step-by-step framework and process for integrating data from multiple sources into a conceptual domain-driven report. This process is not a natural, intuitive one, and thus practitioners are encouraged to practice, get support, and receive guidance and consultation on this process as needed. Key to the process is understanding that every test comes with measurement error, so any single data point from any single test should not be overstated. Each data point from each test should be considered within the context of all other data, from tests, history, and context. Integrating data from different methods and measures is the work of the evaluator, and the process can take quite a bit of time to complete and communicate in writing.

Validity

Emphasis

Careful consideration should be given to the appropriate emphasis of conclusions, particularly when indicating the relative intensity of a client's behavior. General summaries may be given, such as "this client's level of depression is characteristic of inpatient populations," or the relative intensity of certain aspects of a client's disorder may be discussed more specifically. To continue with the example of depression, a clinician may discuss the client's cognitive self-criticisms, degree of slowed behavior, extent of social support, level of social skills, or suicidal potential. In addition to discussing and giving the appropriate degree of emphasis to a client's pathology, his or her psychological strengths need to be compared with his or her relative weaknesses. Furthermore, the report should not discuss areas of minor relevance unless they somehow relate to the purpose of the evaluation. To achieve proper emphasis, the examiner and the referral source must clarify and agree on the purpose of the evaluation. Only after this has been accomplished can the examiner decide whether certain information should be elaborated in depth, briefly mentioned, or deleted.

When clinicians present their conclusions, they should indicate their relative degree of certainty. Is a specific conclusion based on an objective fact, or is the clinician merely presenting a speculation? For example, the statement "John scored in the low average range of intelligence" is an objective fact. However, even in this case, examiners may want to give the standard error of measurement to provide an estimate of the probable range of scores. If only mild supporting data is available or if clinicians are presenting a speculation, phrases such as "it appears . . .," "tends to . . .," "probably . . .," or "likely . . ." (when slightly more certain) should be used. Doing this is especially important when clinicians are attempting to predict a person's behavior, because the predicted behavior has not yet been observed. It may be useful for clinicians to indicate that their predictions cannot be found directly in the tests themselves but rather represent inferences that have been made based on the test data. There should be a clear distinction between what the client did and what he or she anticipates doing. If a statement made in a report is a speculation, it should be clearly indicated that the statement has only a moderate or small degree of certainty. Whenever a speculation is included, it should be relevant to the referral question.

In addition to using language to communicate the degree of certainty around particular conclusions, clinicians should also include some statements about the relative validity of the test findings. For example, a behavioral observation that the client

appeared to be giving genuinely effortful attempts on all tests administered can help the reader understand that the test results are likely a valid representation of the client's best efforts. Alternatively, if obvious fatigue manifested toward the end of a session, when a particular test was administered, this can warn the reader that the test findings are likely an underestimate of the client's functioning. Another way of communicating relative validity of different pieces of data, without stating it outright, is again through emphasis throughout the report. Clinicians should consider, for example, making stronger, more definitive statements based on data that emerged from multiple sources and particularly from more empirically supported tests. Data that emerge from projective techniques, like the Thematic Apperception Test (TAT) or projective drawings, may add nuance to major findings, but these data should be deemphasized somewhat and qualified, as they come from measures that do not have strong empirical validity support. For example, if multiple measures (including the MMPI-2-RF and R-PAS) reflect that the client has depressive ideation, the TAT may suggest that the depressive ideation is predominantly helpless rather than hopeless, worthless, or lonely. Appropriate deemphasizing language for this finding could include: "The TAT suggested that the depressive thoughts he is experiencing *seem to be* mostly about helplessness, not being able to change his current situation or help himself improve."

Improper emphasis can reflect an incorrect interpretation by the examiner, and this misinterpretation is then passed down to the reader. Clinicians sometimes arrive at incorrect conclusions because their personal bias results in selective perception of the data. Thus, clinicians can develop an overly narrow focus with which they overlook potentially relevant data. Personal bias may result from factors such as a restrictive theoretical orientation, incorrect subjective feelings regarding the client, an overemphasis on pathology, or even overconfidence in particular tests. Inaccurate conclusions can also result from attempts to please the referral source or from interpretations based on insufficient data. The reader may also be likely to misinterpret the conclusions if the report is generally overspeculative or if speculations are not specified as such but, rather, are disguised as assertions. Overly assertive speculations may not only lead the reader to develop incorrect conclusions; the report may become overly authoritative and dogmatic, perhaps leading readers to become irritated and skeptical.

Misinterpretations can also result from vague and ambiguously worded sentences that place incorrect or misleading emphasis on a client's behavior. A statement such as "the client lacks social skills" is technically incorrect because the client must have some social skills, although these skills may be inadequate. A more correct description would be to state that the client's social skills are "poorly developed" or "below average." Likewise, a statement such as "the client uses socially inappropriate behavior" is subject to myriad interpretations. It could be rephrased to include more behaviorally oriented descriptions, such as "frequently interrupts" or "would often pursue irrelevant tangents."

The areas, extent, and method of emphasis significantly contribute to the conclusions of a report. However, responsibility for a report's conclusions rests on the clinician. This responsibility should not and cannot be transferred to the tests themselves or to the reader. To take this a step further, decisions made about a person should never be in the hands of tests, which necessarily have some error and may even have questionable validity in certain contexts. Rather, conclusions and decisions regarding

people should always be in the hands of responsible persons. Thus, the style of emphasizing results should reflect this. Phrases such as “test results indicate ...” may give the impression that the examiner is trying to hide behind and transfer responsibility for his or her statements onto the tests. Not only is this not where the responsibility should be, but the reader may develop a lack of confidence in the clinician. If clinicians feel uncertain about a particular area, they should either be clear about this uncertainty or, if they cannot personally stand by the results, exclude the results from the report.

Addressing Diversity

Although not exclusively related to the report, test interpretations and conclusions in the report should be mindful of issues of culture and diversity. Each discussion of individual tests in this book includes a section on use and application of the test with diverse populations. However, more often than not, there will be some cultural aspect of a client that is not perfectly addressed by the tests. While it is often beneficial to address cultural issues explicitly within a report, doing so should accomplish a specific goal. That is, if test interpretations, for example, should be cautiously applied because of cultural concerns, stating this explicitly is beneficial. If there is no direct purpose for discussing cultural issues within the report, doing so can be distracting and even potentially misleading. Even if not addressed explicitly, another professional should be able to read the report and be confident that no culturally inappropriate interpretations or assertions have been made throughout.

Diagnosis and Recommendations

Any reader of an assessment report should be able to follow the logic behind the conclusions, diagnosis, and recommendations. Although diagnosis may or may not be included in the report, if it is, it should correspond logically to the data that emerged from the assessment itself. No diagnosis should be reported that was not supported somehow in the report, even if it is not the main focus of the referral questions. For example, someone in recovery from alcoholism, who has been sober for 20 years, may present for an assessment to aid his or her therapist in deciding how to focus treatment to improve progress. Even though a diagnosis may not be included in such a report, diagnosis may certainly have an impact on the type of treatment recommended. The alcohol dependence may not be the focus of the current assessment, but the diagnosis of alcohol dependence, in remission, should not appear at the end of the report without any supporting information earlier in the report. In this case, it may simply be mentioned in the background information or history sections, self-reported, but the reader should be able to understand where the diagnosis came from. This is true for recommendations as well. The reader should be able to track logically why each recommendation is being made. For less psychologically knowledgeable clients, more explanation may be necessary about the individual recommendations, such as an explanation of a certain type of therapeutic intervention being particularly effective for a certain type of symptom.

Client-Centeredness

Case-Focused Reports

One general style of report to avoid is sometimes referred to as a shotgun report (Tallent, 1992, 1993). A shotgun report provides a wide variety of often-fragmented descriptions in the hope that some useful information can be found within. The shotgun approach is usually vague, stereotyped, and overinclusive. The recommendations for treatment are often neither specific nor practical. The most frequent reason for a shotgun report is a referral question that is too general, vague, and, consequently, poorly understood. In contrast, the case-focused report centers on the specific problems outlined by the referring person. It reveals unique aspects of the client and provides specific, accurate descriptions rather than portraying stereotypes that may also be overly theory linked or test linked. Furthermore, the recommendations for treatment are both specific and practical. The general approach of the case-focused report is not so much *what* is to be known but rather *why* different types of information are important for the purposes of the report.

The creation of a case-focused report involves understanding and applying several basic principles. First, the report should be integrated rather than organized around disparate portions of test data. In other words, the clinician should describe a person rather than merely reporting test data. Second, the recommendations in a case-focused report need to directly relate to what specifically can be done for this client in his or her particular environment. They may apply to areas such as occupational choice, psychotherapy, institutional programs, or additional evaluation. For certain types of referrals, however, especially clients self-referred for psychotherapy, an important goal may be to help them increase their level of personal insight. In these cases, a wider description of the client that includes a number of different topics might be more appropriate than the narrower, problem-solving approach. In addition, there should be a focus on that which differentiates one person from another. Making these differentiations means avoiding discussions of what is average about the client and emphasizing instead what stands out and is unique to this individual. Case-focused reports also frequently deemphasize diagnosis and etiology. There is, rather, a focus on current descriptions of the person that are tied to specific behaviors. In certain cases, especially in a medical setting, the clinician may still need to provide diagnoses in addition to behaviorally oriented descriptions. Another consideration is that a case-focused report should be written with an awareness of the point of view of the intended readers. Appreciating the readers' perspective includes taking into consideration their level of expertise, their theoretical or professional orientation, the decisions they are facing, and the possible interpretations they are likely to make of the information.

Individualization

Related to case-focused reports and integration of test and other data, reports should present data not as single, disparate points of information, but as coherent, narrative, related ideas about the individual person being assessed. It is often tempting to

highlight a single finding from a single test, especially if the clinician has great confidence in that test and that finding. The finding, though, should be contextualized with all the other findings from the assessment, such that it is not overstated. Psychological assessment reports (and the assessors themselves) carry a great deal of authority and power, and clinicians need to ensure that assertions made in the report are specifically about the person being assessed. Related to this, while clinicians can rely on computer interpretations of tests and narrative interpretations of individual test scores like those found in this book, they should avoid copying them verbatim into reports. Most often, the language is not optimal for reports anyway. More important, though, copying such material verbatim necessarily presents individual findings outside of the context of other data and information about the individual being assessed. Even when presenting a generally straightforward idea that has emerged clearly from multiple data sources, it is best practice to individualize the language of that information as it applies specifically to the individual being assessed.

In addition to recommendations being logical conclusions based on the rest of the report (see the “Validity” section above), they should be tailored specifically to the individual being assessed. Recommendations should be clear, specific, and reasonable. For example, though recommending that someone “receive treatment” may be both clear and reasonable, it is not specific enough for the individual to know exactly what next steps to take. Client-centered recommendations must take into consideration how available and accessible resources are for the individual. A clinician may, for example, diagnose a client with borderline personality disorder, and a recommendation for dialectical behavior therapy (DBT) is certainly logical, as well as clear and specific. However, if the client lives in an area where he or she cannot receive such treatment, or if he or she does not have access to dialectical behavior therapy for financial or other reasons, this is not a reasonable recommendation. A client-centered recommendation should take into consideration what is reasonable and realistically doable for the client.

Overall Writing

There are several good resources for helping clinicians with the overall writing, style, and language of assessment reports. First, sample reports like those in this chapter and in Wright (2010) can serve as good models. Further, Groth-Marnat and Davis’s (2014) *Psychological Report Writing Assistant* provides excellent technical writing guidelines and even software to assist in writing reports. Finally, Allyn’s (2012) *Writing to Clients and Referring Professionals about Psychological Assessment Results: A Handbook of Style and Grammar* provides technical assistance on structure, grammar, and specific language to use in reports.

Style

The style, or “flavor,” of a report is influenced primarily by the training and orientation of the examiner. The clinician can choose from four general report-writing approaches: literary, clinical, scientific, and professional (Ownby, 1997; Tallent, 1992, 1993). Each style has unique strengths, and all have a number of liabilities. The literary approach uses everyday language and is creative and often dramatic. Although it can effectively

capture a reader's attention and provide colorful descriptions, it is often imprecise and prone to exaggeration.

The clinical approach focuses on the pathological dimensions of a person. It describes the client's abnormal features, defenses, dynamics involved in maladjustment, and typical reactions to stress. The strength of the clinical approach is that it provides information about areas in need of change and alerts a potential practitioner to likely difficulties during the course of treatment. However, such a report tends to be one-sided and may omit important strengths of the person. The result is likely to be more a description of a "patient" than a person. Such a maladjustment bias is a frequent difficulty in clinical psychology and results in a distorted, unrealistic view of the client. Although most clinical reports should describe a person's problem areas, these problem areas should be given appropriate emphasis in the context of the client's relevant strengths and resources.

The scientific approach to report writing emphasizes normative comparisons, tends to be more academic, and, to a lesser extent, relates to the nature of a client's pathology. The scientific style differs from the other two approaches discussed chiefly in its reference to concepts, theories, and data. It looks at and describes test findings in an objective, factual manner. Thus, there might be frequent references to test data, normative comparisons, probability statements, and cutoff scores to be used for decision making. A scientific approach is likely to discuss the person by addressing different, often isolated, segments of personality. Thus, areas such as a client's cognitive, perceptual, and motivational abilities may be described as discrete and often unrelated functions. Although the scientific approach is objective and factual, it has been criticized for violating the unity of personality. Many readers, particularly those from other disciplines, do not respect or empathize with scientific evaluations and perceive them as cold, distant, and overly objective. Purely data-oriented evaluations have the potential of doing the profession a disservice by reinforcing the view that an assessment is like a laboratory test rather than a professional consultation with a clinician. Furthermore, a focus on factual data may not address the practical decisions the client and referral source are facing.

In actual practice, it is unusual to find a pure example of a literary, clinical, or scientific report. Clinicians generally draw from all three approaches but typically emphasize one. An important part of effective report writing is the ability to evaluate the assets and limitations of each style and to maintain a flexible orientation toward appropriately combining them. In any one report, there may be a need to use creative literary descriptions, elaborate on different pathological dimensions, or provide necessary scientific information. Again, the key is to avoid the pitfalls associated with focusing exclusively on any one of these styles and to emphasize instead their relative strengths.

Ownby (1997) stressed that the most important style to use in report writing is what he refers to as a *professional style*. This style is characterized by short words that are in common usage and that have precise meanings. Grammatically, writers should use a variety of sentence constructions and lengths to maintain the reader's interest. The paragraphs should be short and should focus on a single concept. Similar concepts should be located close to one another in the report. Whereas Hollis and Donna (1979) urged writers to use short words, short sentences, and short paragraphs, the *Publication Manual of the American Psychological Association* (6th ed., 2009) recommended

varying the lengths of sentences and paragraphs. The result should be a report that combines accuracy, clarity, integration, and readability.

Terminology

Several arguments have been made regarding whether to use technical or nontechnical language in psychological reports. It might be argued that technical terminology is precise and economical, increases the credibility of the writer, and can communicate concepts that are impossible to convey through nontechnical language. However, a number of difficulties are often encountered with the use of technical language. One of the more frequent problems involves the varying backgrounds and levels of sophistication of the persons reading the report. The most frequent readers of reports, outside of clients themselves, include teachers, administrators, judges, attorneys, psychiatrists, nonpsychiatric physicians, and social workers (Harvey, 1997, 2006). Thus, most consumers of reports do not have the necessary background to interpret technical terminology accurately. Even psychologists with different theoretical orientations may be apt to misinterpret some of the terms, and psychiatrists may use similar technical terms in slightly different ways than psychologists. Take, for example, the differing uses of *ego* by Freud, Jung, and Erikson. Also, the term *anxiety* might have several different categories of use. Although technical words can undoubtedly be precise, their precision is helpful only in a particular context and with a reader who has the proper background. Generally, reports are rated as more effective when the material is described in clear, basic language (Brenner, 2003; Finn, Moes, & Kaplan, 2001; Groth-Marnat & Davis, 2014; Harvey, 1997, 2006; Ownby, 1990, 1997; Sandy, 1986; Tallent, 1993; Wright, 2010). Even among readers who have the proper background to understand technical terms, many prefer a more straightforward presentation (Ownby, 1990, 1997; Tallent, 1992, 1993). Technical terms also run the danger of becoming nominalisms in which, by merely naming the phenomenon, persons develop an illusory sense of understanding more than is actually the case. Terms such as *immature* or *sadistic* cover a great deal of information because they are so general, but they say nothing about what the person is like when he or she is behaving in these maladaptive ways. They also do not adequately differentiate one person from the next and are frequently ambiguous. Furthermore, technical terms are often used inappropriately (e.g., a person who is sensitive and cautious in interpersonal relationships is labeled *paranoid* or *hypervigilant*, or *compulsive* is used to describe someone who is merely careful, conscientious, and effective in dealing with details).

W. G. Klopfer (1960) provided an excellent and still-relevant rationale for using basic English rather than technical terminology. First, and perhaps most important, the use of basic English allows the examiner, through his or her report, to communicate with and affect a wide audience. This ability is particularly important because the number and variety of persons who read reports is much greater now than even 20 or 30 years ago. Furthermore, basic English is more specific and descriptive of an individual's uniqueness, whereas technical terms tend to deal with generalities. Terms such as *sadomasochistic* and *hostile* do not provide essential information about whether the person is assaultive or suicidal. Finally, the use of basic English generally indicates that the examiner has more in-depth comprehension of the information he or she is dealing with and can communicate this comprehension in a precise, concrete manner. Klopfer

stressed that any description found in a psychological report should be comprehensible to any literate person of at least average intelligence. In contrast, psychologists have been found to write in a more technical and complex level when compared with the average client (Harvey, 1997, 2006). The first four of the next examples show how to translate technical concepts into basic English (Klopfer, 1960):

“Hostility toward the father figure” becomes “the patient is so fearful and suspicious of people in positions of authority that he automatically assumes an aggressive attitude toward them, being sure that swift retaliations will follow. He doesn’t give such people an opportunity to demonstrate their real characteristics since he assumes they are all alike.”

“The patient projects extensively” becomes “the patient has a tendency to attribute to other people feelings and ideas originating within himself regardless of how these other people might feel.”

“The defenses the patient uses are ...” becomes “the methods characteristically employed by the patient for reducing anxiety are ...”

“Empathy” becomes “the patient can understand and sympathize with the feelings of others, since she finds it relatively easy to put herself in their place.”

“The client is hostile and resistant” may be changed to include a behavioral description; “when the client entered the room she stated, ‘My dad said I had to come and that’s the only reason I’m here’” or “later on in the testing she made several comments such as ‘This is a stupid question.’” (pp. 58–60)

The general principle involved in the preceding examples is to translate high-level abstract terms into basic English that provides useful, concrete behavioral descriptions.

Ownby (1990, 1997) recommended combining any conclusion or generalization with specific behaviors or test observations. Recommendations should also be directly linked with relevant behaviors/generalizations, either in the same place or in the recommendations section. For example, instead of saying a client is “depressed,” a writer might state, “The client’s behavior, which included self-criticism and occasional crying, suggested he was depressed, which emerged from the tests as well.” Linking generalizations with clear, concrete descriptions tends to create reports that are perceived to be relatively credible and persuasive (Ownby, 1990, 1997). If this process is followed, descriptions will be less subject to misinterpretation, less ambiguous, and more likely to convey the unique characteristics of the client. Although at times abstract technical terms can be important components of a psychological report, they should be used sparingly and only when clearly appropriate. An essential aspect of whether or not to include technical information is carefully considering the background of the persons who will read the report. Some authors recommend collaborating with the relevant recipients of the report, including the client, so that the final report is descriptive rather than interpretive and the readers are not passive recipients of the “higher” wisdom of the psychologist (Finn, Fischer, & Handler, 2012; Sandy, 1986).

Use of Raw Data

When writing the impressions and interpretation section, a report writer should generally avoid adhering too closely to the raw data. For certain purposes, however, it may

be useful to include raw data or even to describe the tests themselves. Test descriptions allow untrained persons to know specific behaviors the client engaged in rather than merely the final inferences. As a result, consumers of reports rate the inclusion of behavioral referents quite favorably (Finn et al., 2001). For example, a report may include a description such as “Mr. A had an average level of recall for short-term visual information, as indicated by his being able to accurately recall and reproduce five out of a possible nine geometric designs that he had previously worked with for five minutes.” This sentence provides a more behaviorally referenced description than one like “Mr. A had an average level of recall as measured on the Bender–2 Recall.” Another example might be to include a portion of a TAT story (e.g., “At times Mr. A has strong impulses that he finds difficult to control; for example, on a task in which he was asked to make up a story about a picture, he stated ‘... so he took the violin and, without even thinking about it, threw it into the fire and ran’”). These and similar strategies are likely to give the reader a more in-depth, precise, and familiar reference regarding the client’s abilities and personality. However, it is important to ensure that behavioral examples and raw data do not distract the reader, make the findings less clear, or overcrowd the report. Further, actual test items should not be included due to copyright and test security restrictions.

It is crucial to stress that the purpose of providing raw data and behavioral descriptions is to enrich and illustrate the topic and not to enable the reader to follow the clinician’s line of reasoning or document the inferences that have been made. In developing inferences, clinicians must draw on a wide variety of data. They cannot possibly discuss all the patterns, configurations, and relationships they used to come to their conclusions. Any attempt to do so would necessarily be overly detailed, cumbersome, and incomplete. Statements such as “In considering the pattern of elevated Scales 4 and 9 on the MMPI-2, it is safe to conclude...” are unnecessary and rarely contribute to a report’s overall usefulness. In certain types of reports, such as those for legal purposes, it might be helpful to include some raw data, not so much to repeat the thinking process of the clinician but more to substantiate that the inferences are data-based, to provide a point of reference for discussing the results, and to indicate what assessment procedures were used.

FEEDBACK

Although not necessarily directly related to the assessment report, feedback is an important topic in the process of psychological assessment. During the earlier days of psychological assessment, examiners often kept the results of psychological assessments carefully concealed from the client. There was often an underlying belief that the results were too complex and mysterious for the client to understand adequately. In contrast, current practices are to provide the client with clear, direct, and accurate feedback regarding the results of an evaluation (S. J. Ackerman, Hilsenroth, Baity, & Blagys, 2000; Finn, 2007; Finn et al., 2012; Finn & Tonsager, 1997; Lewak & Hogan, 2003; Pope, 1992).

The change toward providing feedback to clients has been motivated by several factors. First, regulations have supported a growing list of consumer rights, including the

right to various types of information. Second, it might be perceived as a violation of trust if the client did not receive feedback regarding the results of testing after he or she had been subjected to several hours of assessment. Even the most secure of clients might easily feel uncomfortable knowing a report with highly personal information, which they have not seen, might be circulated and used by persons in power to make decisions about his or her future. Such practices could understandably result in suspicion and irritation on the part of the public. Third, examiners cannot safely assume that the original referral source will provide feedback to the client. Even if the referral source does provide feedback, there is no guarantee that the information will be provided in an appropriate manner. Thus, the responsibility for providing feedback is ultimately on the clinician. Finally, there is increasing evidence that providing clients with test feedback can result in significant therapeutic benefits (S. J. Ackerman et al., 2000; Finn, 2007; Finn et al., 2012; Finn & Tonsager, 1992; Gass & Brown, 1992).

The extent to which a clinician providing feedback will allow the client to actually read all or portions of the report varies. The rationale for allowing the client actually to read the report is that doing so enables the client to experience the product of assessment in a direct manner. It also enables a practitioner to explain any areas that are unclear. A significant difficulty is that the client might misinterpret various portions of the report, especially IQ scores and diagnosis. For this reason, most clinicians paraphrase and elaborate on selected portions of the report. This method increases the likelihood that clients will readily understand the most important material and will not be overloaded with too much content.

Wright (2010) has characterized feedback sessions with clients as “hybrid” sessions, mixing components of testing (as a researcher/scholar) and therapy (as a clinician). The likelihood of providing effective feedback can be enhanced by following several guidelines. Initially, the rationale for assessments should be explained, and any misconceptions should be corrected. One particularly important misconception is that sometimes clients mistakenly fear that the purpose of assessment is to evaluate their sanity. Practitioners must also select the most essential information to be conveyed to the client. To a large extent, doing this involves clinical judgment. Important considerations include the client’s ego strength, life situation, stability, and receptiveness to different types of material. Typically, three to four general and well-developed areas represent an optimum amount of information. The information that is provided should be integrated carefully into the overall context of the person’s life. This integration might be enhanced by providing concrete behavioral examples, reflecting on aspects of the client’s behavior, referring to relevant aspects of the client’s history, or paraphrasing and expanding on a client’s self-descriptions.

A useful technique is to have the client evaluate the relevance and accuracy of the information. The client might also be asked to give his or her own examples of the trait or pattern of behavior described in the report. Such collaboration with the client helps the clinician to determine how well the client has understood the feedback. Underlying any feedback should be an attempt to provide the information in a clear, intelligible manner. Commonplace language should be used instead of psychological jargon. It is also important to take into account the client’s level of intelligence, education, vocabulary, and level of psychological sophistication. Feedback should be not just a neutral conveyance of data but also a clinical intervention. The information should

provide the client with new perspectives and options and should aid in the client's own problem solving.

One possibility is to prepare a personalized report designed specifically for the client. Doing this forces the practitioner to write in a clear, straightforward style. Such reports are more likely to emphasize adaptation rather than pathology. In addition, clear recommendations tend to be emphasized. The optimal communication style is an informal letter written to and for the client ("I am writing to communicate the findings of our psychological assessment."). There are currently available a number of computerized reports directed toward providing the client with feedback. There also is a trend for additional resources to include interpretations directed toward the client, such as Levak, Siegel, and Nichols's (2011) *Therapeutic Feedback with the MMPI-2: A Positive Psychology Approach*; Finn, Fischer, and Handler's (2012) *Collaborative/Therapeutic Assessment: A Casebook and Guide*; and Finn's (2007) *In Our Client's Shoes: Theory and Techniques of Therapeutic Assessment*. These sources should not be seen as substitutes for a dynamic interaction with a client but as adjuncts for enhancing this process.

FORMAT FOR A PSYCHOLOGICAL REPORT

Although no single, agreed-on format exists, every report should integrate old information as well as provide a new and unique perspective on the person. Old information should include identifying information (name, birth date, etc.), reason for referral, and relevant history. New information should include assessment results and impressions, summary/conclusions, and recommendations. At the top of the report, practitioners should indicate its confidential nature by writing "Confidential Psychological Evaluation." A suggested outline follows.

Name:

Age (date of birth):

Sex:

Ethnicity:

Date of Report:

Name of Examiner:

Referred by:

- I. Referral Question
- II. Evaluation Procedures
- III. Background Information (Relevant History)
- IV. Behavioral Observations
- V. Interpretations and Impressions
- VI. Summary and Recommendations

Although this outline represents a frequently encountered format, there are many variations. Some practitioners prefer to include the client's marital status, occupational status, and handedness (for neuropsychological reports) at the top of the report, along

with the other demographic information. Other practitioners prefer to include test results as a separate section or include additional sections on diagnosis, case formulation, or summary. Sometimes it might be more appropriate to write the report directly to the referral source in a letter format (“Dear Dr. Jones”). The sample reports included later in this chapter have been chosen to demonstrate a variety of different formats in diverse styles and contexts. Each practitioner needs to develop both the format and the style that most effectively meet his or her client’s and referral source’s needs. In addition, different assessment contexts require different styles and areas of focus.

Referral Question

The Referral Question section of the report provides a brief description of the client and a statement of the general reason for conducting the evaluation. In particular, this section should include a brief description of the nature of the problem. If this section is adequately completed, it should give an initial focus to the report by orienting the reader to what follows and to the types of issues that are addressed. This section should begin with a brief, orienting sentence that includes essential information about the client (“Mr. Smith is a 35-year-old, European American, married male with a high school education who presents with complaints of depression and anxiety”). Such a sentence clearly and succinctly introduces the client. A prerequisite for this section is that the clinician has developed an adequate clarification of the referral question. The purpose of testing should be stated in a precise and problem-oriented manner. Thus, phrases like “The client was referred for a psychological evaluation” or “as a requirement for a class project” are inadequate because they lack focus and precision. It is helpful to include both the specific purpose of the evaluation and the decisions facing the referral source.

Examples of possible reasons for referral include:

- Intellectual evaluation: routine, intellectual disability, giftedness
- Differential diagnosis, such as the relative presence of psychological difficulties (e.g., memory problems caused by depression) versus organic impairment (e.g., memory problems because of the early stages of Alzheimer’s disease)
- Assessment of the nature and extent of brain damage
- Evaluation as a component of, and to provide recommendations for, vocational counseling
- Evaluation of appropriateness for, possible difficulties encountered in, and optimal approach to psychotherapy
- Personal insight regarding difficulties with interpersonal relationships
- Evaluation as an aid in client placement

These reasons represent general referral questions that, in actual situations, would require further clarification, especially regarding the decisions facing the referral source (see Armengol, 2001). The key should be to find out what the referring person really wants from the report. Doing this may require reading beneath the surface of the referral question(s) and articulating possible hidden agendas and placing the referral

question into a wider context than the presenting problem. In some cases, it may be necessary to educate the referral source regarding the strengths as well as the limitations of psychological testing, which may even lead the practitioner to recommend that the person not be tested. An effective referral question should accurately describe the client's and the referral source's current problems.

After the referral questions have been clarified and outlined, they can be addressed throughout the rest of the report. It is usually helpful to succinctly reiterate and summarize the answers to the referral questions toward the end of the report. A useful strategy is to number each of the referral questions listed in the Referral Question section of the report and follow this up with succinct answers to each question in the Summary and Recommendations section. The numbers of the questions and the answers that correspond to them should be identical. Such a method is user friendly; provides clear answers to the questions; and allows for symmetry, integration, and closure to the report.

Evaluation Procedures

The report section that deals with evaluation procedures simply lists the tests and other evaluation procedures used but does not include the actual test results. Usually full test names are included along with their abbreviations. Later in the report, the abbreviations can be used, but the initial inclusion of the entire name provides a reference for readers who may not be familiar with test abbreviations. For legal evaluations or other occasions for which precise details of administration are essential, it is important to include the date on which different tests were administered and the length of time required to complete each one. For most routine evaluations, however, this degree of detail is not recommended. It may also be important to include whether a clinical interview or mental status examination was given and, if so, the degree of interview structure and the amount of time required for the interview or examination. Evaluation procedures may not necessarily be restricted to testing and interviews with the client. Often evaluation includes a review of relevant records, such as medical reports, nursing notes, military records, police records, previous psychological or psychiatric reports, or educational records. Additional material might come from interviews with individuals such as spouses, children, parents, friends, employers, physicians, lawyers, social workers, or teachers. Observations may be included as well, such as observing a child in his or her school environment. If any of these sources is used, the dates and, if relevant, who wrote the material should be included. This section might end with a statement summarizing the total time required for the evaluation.

Background Information (also Referred to as Relevant History)

The write-up of a client's background information should include aspects of the person's history that are relevant to the problem the person is confronting and to the interpretation of the test results. The history, along with the referral question, should also place the problem and the test results into the proper context. In accomplishing these goals, the clinician does not need to include a long, involved chronology with a large number of details, but rather should be as succinct as possible. Some practitioners

even urge that the background information section be kept to one concise paragraph, particularly in medical settings, where there is considerable emphasis on conciseness. In selecting which areas to include and which to exclude, a clinician must continually evaluate these areas in relationship to the overall purpose of the report. It is difficult to specify precise rules because each individual is different. Furthermore, each clinician's own personal and theoretical orientation alters the types of information he or she feels are significant. Whereas one clinician may primarily describe interpersonal relationships, another may focus on intrapsychic variables, birth order, early childhood events, or details about the client's current situation and environment. The key is to maintain a flexible orientation so that the reader is aware of the most significant elements in the client's life. In general, the end product should include a good history of the problem along with areas such as important life events, family dynamics, work history, personal interests, medical history, daily activities, and past and present interpersonal relationships (see Table 3.1 in Chapter 3).

When describing a client's background, it is important to specify where the information came from ("The client reported that ..."). This is particularly essential when there may be some question regarding the truth of the client's self-reports or when the history has been obtained from multiple sources.

Usually a history begins with a brief summary of the client's general background. This can be followed by sections describing family background, personal history, medical history, history of the problem, and current life situation.

The extent to which a clinician decides to pursue and discuss a client's family background is subject to a great degree of variability. The primary purpose of such information is to help determine causal factors, what variables might help maintain relevant behaviors, and the extent to which the family should be used as either a focus of systemic intervention or as social support. At a minimum, a brief description of the client's parents is often warranted; this description may include whether they are separated/divorced and alive/deceased and their socioeconomic level, occupation, cultural background, and health status. Sometimes it is important to include information about the emotional and medical backgrounds of parents and close relatives, because certain disorders occur with greater frequency in some families than in the overall population. A description of the general atmosphere of the family is often helpful, including the client's characteristic feelings toward family members and perceptions of their relationships with each other. Descriptions of common family activities and whether the family lived in an urban or a rural environment might also be included. If one or both parents died while the client was young, the clinician can still discuss the speculations the client has about his or her parent(s) and can describe the significant persons for the client as he or she was growing up.

The client's personal history can include information from infancy, early childhood, adolescence, and adulthood. Each stage has typical areas to investigate and problems to be aware of. The information from infancy usually either represents vague recollections or is secondhand information derived from parents or relatives. Thus, it may be subject to a great deal of exaggeration, selective omission, or fabrication. If possible, it may be helpful to have details verified by additional sources, such as through direct questioning of parents or examination of medical records. The degree of contact with parents, family atmosphere, and developmental milestones may all be important areas

to discuss. Often it is important to include a client's early medical history since physical and psychological difficulties can be closely linked. The most significant tasks during childhood are the development of peer relationships and adjustment to school. What was the quality of the client's early friendships? How much time did the client spend with others? Were there any fights or rebellious acting out? Was the client basically a loner, or did he or she have a large number of friends? Did the person join clubs and have group activities, hobbies, or extracurricular interests? In the academic area, it may be of interest to note the client's usual grades, best or worst subjects, and whether the client skipped or repeated grades. Furthermore, what was his or her relationship with parents, and did the parents restrict activities or allow relative freedom? During adolescent years, clients typically face further academic, psychological, and social adjustments to high school. Of particular importance are their reactions to puberty and early romantic and sexual relationships. Did they have difficulties with sex role identity, abuse drugs or alcohol, or rebel against authority figures? The adult years center around occupational adjustment and establishing marital and family relationships. During early adulthood, what were clients' feelings and aspirations regarding romantic relationships or marriage? What were their career goals? Did they effectively establish independence from parents? As adulthood progressed, were there any significant changes in the quality of their close relationships, employment, or expression of sexuality? What activities did they engage in during their leisure time? As clients age, they face challenges in adapting to their declining abilities and limitations, and developing a meaningful view of their lives. It is important to note that much of this may be assessed in an interview (see Chapter 3), but not all of it will be relevant to the referral questions or test interpretations, and much will thus be left out of a report.

Although the personal history can help place the problem in its proper context and explain certain causative factors, it is usually essential to spend some time focusing directly on the problem itself. Of particular importance are the initial onset and the nature of the symptoms. From the time the client first noticed these symptoms, have there been any changes in frequency, intensity, or expression? If a formal diagnosis will be made, it is particularly important to have a clear description of symptom patterns to substantiate such a diagnosis. It might also be important to determine whether there were any previous attempts at treatment, and, if so, the outcome. In some reports, the history of the problem is the longest and most important part of the history section.

The family and personal histories usually reveal information relating to the predisposing cause of a client's difficulties, whereas the history of the problem often provides an elaboration of the precipitating and reinforcing causes. To complete this picture, the clinician also has to develop a sense of the factors currently reinforcing the problem. Doing this requires information relating to the client's life situation. Significant areas may be the client's life stresses, including changes that he or she is confronting. In addition, what are the nature of and resources provided by his or her family, social, and work relationships? Finally, it is important to understand the alternatives and decisions that the client is facing.

Sometimes an evaluation needs to assess the possible presence and nature of organic impairment. In many cases, the history is of even greater significance than test results; and, often, the most valuable information a psychologist can provide to a referring

medical practitioner is a thorough history. Thus, the history needs to be complete and must address a number of areas that are not ordinarily covered in personality evaluations. Several interview aids have been commercially developed to help ensure that most relevant areas are covered (see Chapter 12). If the person reports having had a head injury, it is important to note the length of time the client was unconscious (if at all), whether he or she actually remembers getting hit, the last memory before the injury, and the first thing he or she clearly remembers following the injury. In all neuropsychological assessments, a crucial area is to establish the person's premorbid level of functioning. To do so, the clinician may have to obtain information on his or her grade point average in high school or college, send for any relevant records (e.g., previous IQ results), and determine previous highest level of employment and personal interests or hobbies. Often it may be necessary to verify the client's previous level of functioning with outside sources, such as parents, spouse, children, or employers. In determining the probable cause of brain impairment, it may be difficult to rule out other possibilities, such as exposure to toxic substances, strokes, high fevers, or other episodes of head trauma. Areas of current functioning that need to be addressed might include memory problems, word-finding difficulties, weakness on one side of the body, alterations in gait, loss of consciousness, and unusual sensations. Previous assessments with computed tomography/magnetic resonance imaging/functional magnetic resonance imaging (CT/MRI/fMRI) scans, electroencephalograms (EEGs), or neurological physical exams would also be important to obtain. Even though these medical records might be able to identify the site and size of a lesion, it is still the work of the psychologist to describe how the person is functioning as a result of the lesion. It might also be important to obtain current or past information regarding drug intake (legal and illegal), especially recent alterations in prescriptions, because these might affect psychological functioning. The interview data and neuropsychological test results from a psychologist should ideally be combined with and complement medical records, such as CT scans and neurological exams. Although the preceding topics are by no means exhaustive, they represent some of the more important areas to consider when taking a history related to possible neuropsychological deficit.

Although the quantity of such information may seem immense, the history format described here is only a general guideline. At times, it may be appropriate to ignore many of the areas mentioned earlier and focus on others. In condensing the client's history into the report, it is important to avoid including superfluous material and continually question whether the information obtained is relevant to the general purpose of the report. Thus, the History section of the report should include all relevant information but should not be overly inclusive.

Behavioral Observations

A description of the client's behaviors can provide insight into his or her problem and may be a significant source of data to evaluate test validity and confirm, modify, or question the test-related interpretations. These observations can be related to a client's appearance, general behavioral observations, or examiner–client interaction. Descriptions should generally be tied to specific behaviors and should not represent a

clinician's inferences. For example, instead of making the inference that the client was "depressed," it is preferable to state that "her speech was slow, and she frequently made self-critical statements, such as 'I'm not smart enough to get that one right.'"

Relevant behavioral observations made during the interview include physical appearance, behavior toward the task and examiner, and degree of cooperativeness. A description of the client's physical appearance should focus on any unusual features relating to facial expressions, clothes, body type, mannerisms, and movements. It is especially important to note any contradictions, such as a 14-year-old boy who acts more like an 18-year-old or a person who appears dirty and disheveled but has an excellent vocabulary and high level of verbal fluency. The behaviors the client expresses toward the test material and the examiner often provide a significant source of information. These may include behaviors that reflect the person's level of affect, manifest anxiety, presence of depression, or degree of hostility. The client's role may be as an active participant or generally passive and submissive; he or she may be very much concerned with his or her performance or relatively indifferent. The client's method of problem solving is often a crucial area to note, and it may range from careful and methodical to impulsive and disorganized. It is also important to pay attention to any unusual verbalizations that the client makes about the test material. The level of cooperation expressed by the client, in addition to behaviors that reflect effortful attempts on tests, should be a factor in assessing the validity of the test results. Level of cooperation is especially important for intelligence and ability tests, because a prerequisite is that the client be alert and attentive and put forth his or her best effort. It may also be important to note events before testing, such as situational crises, previous night's sleep, or use of medication. If there are situational factors that may modify or bring into question the tests' validity, they should be noted with statements such as "The test results should be viewed with caution because..." or "The degree of maladjustment indicated on the test scores may represent an exaggeration of the client's usual level of functioning due to conditions surrounding the test administration." Often the most important way to determine test validity in relationship to the client is through a careful look at the client's behaviors relating to the tests and his or her life situation before testing.

Behavioral observations usually should be kept concise, specific, and relevant. If a description does not allow for some insight about the person or demonstrate his or her uniqueness, it should not be included. Thus, if a behavior is normal or average, it is usually not important to discuss other than to briefly mention that the person had, for example, an average level of cooperation, alertness, or anxiety. The focus, then, should be on those client behaviors that create a unique impression. This section usually should not exceed one paragraph. However, in some instances, there may be considerable relevant information that would require two or three paragraphs. The relative importance of this section in relation to the overall report can be quite varied. Sometimes behavioral observations can be almost as important as the test results; at other times, the description might consist of a few minor observations.

Clinicians who prefer behavioral assessment procedures might wish to emphasize the Behavioral Observation section by providing more in-depth descriptions of relevant antecedents. In addition, consequent events surrounding the problem behavior itself might be evaluated in relation to their onset, duration, frequency, and intensity. Specific

strategies of behavioral assessment include narrative descriptions, interval recording, event recording, ratings recordings, and self-report inventories (see Chapter 4).

Some examiners may wish to summarize information from a mental status evaluation in the Behavioral Observations section. In these cases, there is necessarily a movement away from concrete descriptions of behaviors to inferences about these behaviors. For example, a clinician may infer, based on behavioral observations, that the client was oriented to time and place. Additional categories might include verbalizations, psychomotor activity, affect, thought processes/contents, and insight/judgment (see section titled “Mental Status Evaluation” in Chapter 3).

Another exception to adhering closely to concrete behavioral descriptions is that, at the end of the Behavioral Observations section, it is customary and appropriate to include a statement indicating the validity of the assessment procedures. For example, the statement might say something like: “Given the consistency and detail of the client’s responses, the client’s high level of motivation, and validity indicators on the MMPI-2-RF, the assessment appears to be an accurate assessment of her current level of functioning.”

Interpretation and Impressions

For many reports, it may not be necessary to list test scores. Some practitioners do not give actual test scores because they might be misinterpreted and give the impression that the report is too data/test oriented. However, it is often recommended that, at some point, test scores be included, especially in legal reports or when professionals who are knowledgeable about testing will read the report. One option is to include test scores in an appendix. This method has the advantage of removing potentially distracting technical detail from the narrative portion of the report.

If actual test scores are included, standard (rather than raw) scores should be the mode of presentation. Referral sources have consistently indicated that percentiles are preferred over other types of standard scores (Finn et al., 2001). Because various tests use somewhat different types of standard scores, it is recommended that each set of test scores include both the standard score and percentiles. Clinicians may also wish to indicate the relative magnitude of the relevant scores (“Very high,” “High,” etc.) or whether the scores exceed some clinically meaningful cutoff.

If presented in tables or an appendix, intelligence test scores are traditionally listed first and, for the Wechsler scales, should include the Full Scale IQ score, index scores, and subtest scaled scores. Subtests that have been found to be significant strengths should be indicated with an “S” next to the subtest score, and significant weaknesses should be indicated with a “W.” This listing is often followed by other cognitive test results, such as the Bender–2 or Wechsler Memory Scale–IV. Bender–2 results can simply be summarized by a statement such as: “Empirically not in the organic range, although there were difficulties organizing the designs and frequent erasures.” MMPI-2/MMPI-A/MMPI-2-RF results often are listed in the order in which they appear on the profile sheet. Objective personality tests (MMPI-2/MMPI-A/MMPI-2-RF, MCMI-IV, PAI, NEO–PI–R) should always be referred to by their standardized (usually *T*) scores and not their raw scores. Whereas it is fairly straightforward to list the objective and intelligence test scores, it is considerably more difficult

to adequately describe the scores on projective tests. The Rorschach summary sheet can be included, but the results from projective drawings and the TTAT are usually omitted. Should a clinician wish to summarize projective drawings, a brief statement is usually sufficient, such as “Human figure drawings were miniaturized and immature, with the inclusion of two transparencies.” Likewise, TAT “scores” can be summarized by a brief statement of the most common themes encountered in the stories.

Whether or not test results are included, the Interpretation and Impressions section (sometimes simply called the Impressions section) can be considered the main body of the report. In this section, the main findings of the evaluation are presented in the form of integrated hypotheses. The areas discussed and the style of presentation vary according to the personal orientation of the clinician, the purpose of testing, the individual being tested, and the types of tests administered. As emphasized previously, assessment data should be organized according to different integrated topics. In contrast, a test-by-test presentation is strongly discouraged. To organize the information from an assessment, W. G. Klopfer (1960) recommended using a grid with the topics for consideration in the left column with the assessment results in the top row. This format enables the practitioner to extract essential findings from the data and list them in the appropriate box where the topic and the method of assessment intersect. Wright (2010) further described how to use a grid method to evaluate common, related, or conflicting information across methods of assessment. When actually writing the Interpretation and Impressions section of the report, the clinician can review all findings in a particular topic and summarize them in the report. An example of such a grid is given in Table 15.2. The list of assessment methods is dependent on which tests the examiner administered, but the topics can be chosen and arranged according to areas the clinician would like to focus on.

All inferences made in the Interpretation and Impressions section should be based on an integration of the test data, behavioral observations, relevant history, and additional available data. The conclusions and discussion may relate to areas such

Table 15.2 Sample Grid of Assessment Domains by Tests Administered

Topics	Evaluation Procedures					
	Interview	WAIS-IV	MMPI-2-RF	BDI–II	R-PAS	Behavioral observations
Cognitive functioning						
Personality						
Emotional functioning						
Interpersonal relations						
Self-evaluation						
Coping style						
Client strengths						
Diagnostic impressions						

Source: Adapted from *The Psychological Report* by W. G. Klopfer, 1960, New York, NY: Grune & Stratton.

as the client's cognitive strengths and weaknesses, emotional difficulties, coping style, self-concept, dynamics behind the presenting problem, interpersonal relationships, or client strengths. A client's intellectual abilities often provide a general frame of reference for a variety of personality variables. For this reason, a discussion of the client's intellectual abilities usually occurs first. Although this should include a general estimate of the person's intelligence as indicated by IQ test scores, it is also important to provide a discussion of more specific abilities. This discussion may include an analysis of areas such as memory, problem solving, abstract reasoning, concentration, and fund of information. If the report will be read *only* by persons who are familiar with test theory, it may be sufficient to include IQ scores without an explanation of their normative significance. In most reports, it is helpful to include the IQ test scores as well as the percentile ranks and general qualitative intellectual classification (high average, superior, etc.; see Table 5.5). Some examiners may even prefer to omit the actual IQ test scores in favor of including only percentile rank and general classification. This method can be useful in cases in which persons reading the report might be likely to misunderstand or misinterpret unexplained IQ test scores. After a general estimate of intelligence has been made, it should, whenever possible, be followed by a discussion of the client's intellectual strengths and weaknesses. This discussion may involve elaborating on the meaning of the difference between index scores or describing subtest scatter. In addition, it can be useful to compare the client's potential level of functioning with his or her actual performance. If there is a wide discrepancy between these two, reasons for this discrepancy should be offered. For example, the client may be underachieving because of anxiety, low motivation, emotional interference, or perceptual processing difficulties. Practitioners may also wish to discuss additional noncognitive areas of intellectual assessment, which might include the extent to which the person prefers to achieve through independent activities versus a structured environment, the level of motivation, or the relative intellectual efficiency or hardiness. Cognitive assessments in psychiatric contexts might include any bizarre associations, degree to which the person's thinking is organized, or how concrete or abstract his or her thought processes are. Whereas a discussion of intellectual abilities is relatively clear and straightforward, the next sections are frequently more difficult to produce. There are an extremely wide number of possibilities to choose from, many of which are listed in Table 15.1. Some practitioners recommend including set topics, which typically include the client's level of cognitive functioning, emotional functioning (affect and mood), self-concept, and interpersonal relationships. A neuropsychological evaluation might divide the impression and interpretation into areas such as memory, language functions, executive abilities, awareness of deficits, sensory/perceptual functions, and personality (Groth-Marnat, 2000a; Hebben & Milberg, 2002). One rationale for not having a preset list of topics to discuss is that the topics should be based primarily on the referral question. Not having a preset list of topics allows the practitioner to flexibly organize the information based on the context of the report and the needs of the referral source and client. If the referral question is clearly focused on a specific problem, it may be necessary to elaborate on only two or three topics. A referral question that is more general may require a wider approach in which six or more areas are discussed. In general, flexibility in topics can allow for a more client-focused approach to case formulation (see "Case Formulation" section in Chapter 14).

Some additional common and important topics are the client's level of psychopathology, dependency, hostility, sexuality, interpersonal relationships, diagnosis, and behavioral predictions. A client's level of psychopathology refers to the relative severity of the disturbances he or she is experiencing. It is important to distinguish whether the results are characteristic of non-clinical populations, outpatients, or inpatients, and whether the difficulties are chronic or more a reaction to current life stresses. Does the client use coping behaviors that are adaptive or those that are maladaptive and self-defeating? Within the area of ideation, are there persistent thoughts, delusions, hallucinations, loose associations, blocking of ideas, perseveration, or illogical thoughts? It may also be important to assess the adequacy of the client's judgments and relative degree of insight. Can the person effectively make plans, understand the impact he or she has on others, and judge the appropriateness of his or her behavior? To assess the likelihood of successful therapy, it is especially important to assess the client's level of insight. Doing this includes assessing the person's ability to think psychologically, awareness of his or her own changing feelings, understanding of the behaviors of others, and ability to conceptualize and discuss relevant insights.

Discussing clients' characteristic patterns and roles in interpersonal relationships can also be extremely useful. Often these can be discussed in relation to the dimensions of submissiveness/dominance and love/hate, or in relation to the extent to which they orient themselves around the need to be included, control others, or seek affection. Is their style of communicating typically guarded, or is it open and self-disclosing to the extent that they can discuss areas such as painful feelings and fears? Can they deal with the specifics of a situation, or are they usually vague and general? Do they usually appear assertive and direct or passive and indirect? Finally, it is often important to determine the extent to which they are perceptive about interpersonal relationships and their typical approaches toward resolving conflict.

It may also be appropriate to include descriptions of vocational goals and aptitudes. This information often is quite important in educational reports, especially for students with special educational needs, such as those with disabilities. Many of the tests covered in this text can help in assessing a person's strengths and weaknesses, but practitioners may also need to include further assessment devices, such as the Self-Directed Search, the Strong Interest Inventory, or the Kuder Occupational Interest Survey (see Prince & Heisser, 2000).

A frequent consideration is whether the client's difficulties will continue or, if currently absent, recur. If the client's future prospects are poor, a statement of the rationale for this conclusion should be given. For example, if a client has a strong need to appear hypernormal, with poor insight, and a high level of defensiveness, the clinician should explain these as the reason he or she predicts that response to treatment will be poor. Likewise, favorable predictions should include a summary of the client's assets and resources, such as psychological-mindedness, motivation to change, and social supports. If difficulties are likely to be encountered during the course of treatment, the nature and intensity of these difficulties should be discussed. The prediction of suicidal potential, assaultive behavior, child abuse, or criminal behavior is essential in certain types of reports. Often the tests themselves are not useful in predicting these behaviors. For example, one of the best ways of predicting suicidal potential is to evaluate the client's history, current environment, personal resources, and degree of suicide intent.

(Bernert, Hom, & Roberts, 2014; Simon & Hales, 2012). However, research indicates that many predictions of behavior, such as dangerousness, are subject to error (Fowler, 2012; M. Yang, Wong, & Coid, 2010). Making long-term predictions is especially likely to produce a high rate of error. Clinicians thus should exercise appropriate caution in making predictions and not exceed the bounds of reasonable certainty.

Sometimes clinicians may wish to include a separate section on diagnosis. However, whether to include a diagnosis, whether from the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric Association, 2013) or *International Classification of Disorders (ICD-10)* (World Health Organization, 1990), has been an area of some controversy. Some clinicians feel that labels should be avoided because they may result in self-fulfilling prophecies, be overly reductionistic, and allow clients to avoid responsibility for their own behavior. Other objections to diagnosis stem from researchers who feel that many of the terms are not scientifically valid (Beutler & Malik, 2002; Rosenhan, 1973) and are not particularly useful in planning interventions (Beutler & Malik, 2002; Groth-Marnat, Roberts, & Beutler, 2001; Houts, 2002). If a clinician does decide to give a diagnosis, he or she must first have a clear operational knowledge of the diagnostic terms. He or she should also include the client's premorbid level of adjustment and the severity and frequency of the disturbance. Instruments such as the Structured Clinical Interview for the *DSM-5* (SCID-5; First, Williams, Karg, & Spitzer, 2015), Structured Interview for *DSM-IV* Personality (SIDP-IV; Pfohl, Blum, & Zimmerman, 1997), or Anxiety Disorders Interview Schedule (ADIS-IV; DiNardo, Brown, & Barlow, 1994) might help to increase the reliability of diagnosis. It is also important to include the possible causes of the disorder. A discussion of causes should not be simplistic and one-dimensional, but rather should appreciate the complexity of causative factors. Thus, causes may be described from the perspective of primary, predisposing, precipitating, and reinforcing factors. Clinicians may also discuss the relative significance of biological, psychological, and sociocultural variables.

Inclusion of client strengths is becoming increasingly prevalent and relevant. Describing a series of client strengths helps to balance out the typically negative features that occur in many if not most psychological reports (C. R. Snyder, Ritschel, Rand, & Berg, 2006). Not only can the inclusion of strengths provide more balance, but when reports are read by clients, seeing a listing of their strengths can enhance rapport between the examinee and the mental health profession in general. There is also increasing research evidence that tailoring interventions toward a person's unique pattern of strengths can help improve outcome (M. E. P. Seligman, Steen, Park, & Peterson, 2005).

Summary and Recommendations

The purpose of the Summary subsection is to restate succinctly the primary findings and conclusions of the report. To do so, the practitioner must select only the most important issues and be careful not to overwhelm the reader with needless details. As emphasized previously (see the "Referral Question" section), a useful strategy in the Summary section is to provide brief bulleted/numbered answers to each of the referral questions. As stated earlier, the Summary section should be written as if it is the only section the audience will read; it should be able to stand on its own.

The ultimate practical purpose of the report is contained in the recommendations because they suggest what steps can be taken to solve problems. Such recommendations should be clear, specific, practical, and obtainable and should relate directly to the purpose of the report. The best reports are those that help the referral sources and/or the clients solve the problems they are facing (Armengol, Moes, Penney, & Sapienza, 2001; Brenner, 2003; Finn et al., 2001; Groth-Marnat & Davis, 2014; Ownby, 1997; Tallent, 1993). To achieve this report-writing goal, the clinician must clearly understand the problem, the best alternatives for remediation, and the resources available in the community. One practical implication is that clinicians can improve their reports by becoming as familiar as possible with the uses to which their reports will be applied. An effective report must answer the referral questions and have decisional value. After these factors have been carefully considered, recommendations can be developed.

Chapter 14 elaborated on guidelines for psychotherapy with a focus on selecting the optimum intervention procedures along with considerations for enhancing the therapeutic relationship and decisions related to case management. Often cases will require a wider variety of recommendations than this, especially in forensic, medical, academic, or rehabilitation settings. These recommendations may include treatment options, placement decisions, further evaluation, altering the client's environment, use of self-help resources, and miscellaneous considerations. Clinicians can consult Table 15.3 to see if there might be additional recommendations they would like to include in their report, as discussed in Chapter 14.

Self-help resources have become a particularly well-developed, useful, cost-effective option. Depending on the type of problem and characteristics of the person, research supports that many self-help programs are nearly as effective as formal mental health treatment (Norcross, 2006). Norcross et al.'s (2013) *Self-Help That Works: Resources to Improve Emotional Health and Strengthen Relationships* lists a wide variety of resources, rates their quality, and organizes them according to books, autobiographies, films, Internet/online resources, and national support groups. There are also a wide variety of computer-assisted psychotherapy programs (Cucciare & Weingardt, 2009;

Table 15.3 Types of Treatment Recommendations

Recommendation type	Examples
Treatment	Psychotherapy, speech therapy, occupational therapy, medication, meditation, mediation
Placement	Special education, nursing home, 24/7 observation, joint custody, inpatient care
Further evaluation	Reevaluation with selected portions of current tests, physical exam, CT/MRI
Alteration in environment	Medication alarm, internal/external reminders
Self-help	Self-help books, films, websites, support groups; computer-aided interventions
Miscellaneous	Revoke driver's license, wear Medialert bracelet, probation, homework (gratitude letter, practice self-statements)

I. M. Marks, Cavanagh, & Gega, 2007). Self-help resources can be used either in conjunction with formal mental health treatment or as stand-alone interventions.

One clear finding is that reports typically are rated most useful when their recommendations are highly specific rather than general (Armengol, 2001; Finn et al., 2001; Ownby, 1990, 1997; G. W. White, Nielsen, & Prus, 1984). Thus, a recommendation that states “The client should begin psychotherapy” is not as useful as a statement of the need for “individual therapy focusing on the following areas: increased assertiveness, relaxation techniques for reducing anxiety, and increased awareness of the self-defeating patterns he creates in relationships.” Likewise, a recommendation for “special education” can be improved by expanding it to “special education placement along with services 2 hours a day, emphasizing exercises in auditory sequencing and increasing immediate recall for verbally relevant information.” However, caution should be exercised when providing specific recommendations in some contexts because some health professionals may feel that developing treatment recommendations is primarily their responsibility or perhaps should be made by the overall treatment team. After the report has been submitted, continued contact should be made with the readers(s) to make sure the report has not been filed and forgotten. Even the best report is not useful unless the recommendations are practical, obtainable, and actually put into action (Geffken, Keeley, Kellison, Storch, & Rodrigue, 2006).

SAMPLE REPORTS

The sample reports in this section are from the more common settings in which clinicians work and consult. The dimensions in which the reports vary are:

- Format;
- referral question;
- extent to which history rather than test data is emphasized;
- types of tests used; and
- degree to which they include a variety of descriptions versus being case-focused with a relatively limited range of topics.

In each setting, specific questions have been presented, along with decisions that must be made related to the client. The different reports illustrate how the clinician has integrated the test data, client’s history, and behavioral observations to handle these questions. The reports were selected to illustrate a wide diversity in format, length, type of setting, referral question, and type of tests used.

The first report was written in a psychiatric setting and was intended for use by mental health professionals. As a result, some technical language is used, primarily in the form of a formal *DSM-5* diagnosis. A further aspect to note is that the evaluation of the patient’s cognitive functioning was not based on formal testing but rather on behavioral observations and mental status. The major feature of the report is the extensive development of a detailed treatment plan for psychotherapeutic intervention. This plan was developed based on the Systematic Treatment Selection model detailed

in Chapter 14. The recommendations are eclectic in orientation and assume that the treating practitioner can effectively use a number of techniques from a variety of theoretical orientations. Another feature of the report is the absence of psychological test data. Specifically, there is no Test Results section because some, if not many, practitioners believe that the inclusion of detailed test results is both unnecessary and results in cluttering up the report with distracting detail. It is rather assumed that the referral source is most interested in the integration of the overall assessment along with the relevant recommendations.

The second report, written in a legal context, focuses on addressing issues related to comprehension of Miranda warnings, ability to partake in police interrogation, and the possibility of a false confession. As a result, the report relies on measures of cognitive abilities and level of achievement. The referral questions are clear and focused. They are numbered in the Referral Question section and then, using the corresponding numbers, are answered in the Summary section. In most reports, it is neither recommended nor necessary to include citations/references. However, they have been included in this report to support the credibility of the assertions. The argument was that, given available research, the client had an optimal set of characteristics for him to have been unable both to comprehend his Miranda rights and to knowingly and intelligently participate in his police interrogation.

The third sample is from an educational context; the client (a 12-year-old White male) was experiencing difficulties around his academic performance, as well as some behavioral and social difficulties earlier in the year that seemed to have subsided. This report is an important addition not only because it includes an adolescent but also because it demonstrates how the examiner handled complex cognitive and academic achievement findings. The report highlights the need for the evaluator to have a thorough understanding of the possible diagnostic outcomes, as the cognitive findings did not support an obvious, clear diagnosis. Additionally, the emotional section includes a case formulation using the Complex Model as presented in Chapter 14, and the Summary section works to integrate all the cognitive and emotional information together. Because the referral questions were simply diagnosis and treatment recommendations, rather than numbering these and making the recommendations correspond, the diagnosis was given its own section and the treatment plan was part of the Summary and Recommendations. In most cases, it is important to keep the recommendations succinct. This report provides quite a long and comprehensive set of recommendations; however, they are organized according to various persons who will be involved in working with him. For example, there is a set of recommendations for the school (around accommodations), a separate set of recommendations for his learning coach, and so on. Although a long and detailed set of recommendations can sometimes be intimidating to clients, parents, and others who receive reports, clearly separating out the intended audience can help each member of the broader “treatment team” make specific, assessment-driven decisions.

The final report is in the form of a letter written directly to the client. The language is user-friendly, empathic, and sometimes uses the client’s own phrases. It is also clearly organized answering the referral questions. Such a report represents a growing trend toward integrating assessment with therapy. It also reflects a highly collaborative and egalitarian approach to the assessment process.

Psychiatric Setting

NAME: A. G.

DATE OF BIRTH: 5/30/1943

DATES OF EXAMINATION: 12/12; 12/13; 12/21/2013

SEX: Female

ETHNICITY: European American

REFERRED BY: Dr. M.

REFERRAL QUESTION

A. G. is a 70-year-old, divorced female with 13 years of education who reported anxiety and episodes of dissociation. The patient reported that she has experienced agoraphobia in the past but now clarifies that her problem is one of motivation rather than panic. She asserted that she is currently not immobilized nor is she extremely anxious when she travels. She acknowledged an “underlying apprehension” that arises when she is scheduled to leave home, however, resulting in her putting off her departure as long as possible. A. G. was referred by her psychiatrist, Dr. M., who has requested an evaluation to clarify the following questions:

1. What is the nature and dynamics of the problem?
2. What are the client’s strengths?
3. What is an optimal treatment plan?

EVALUATION PROCEDURES

Life History Questionnaire (12/12/2013).

Clinical Interview (12/12/2013).

Structured Clinical Interview for DSM-IV (SCID; 12/22/2013).

State-Trait Anxiety Inventory (STAI; 12/13/2013).

Millon Clinical Multiaxial Inventory—III (MCMI-III; 12/13/2013).

Personal Attitude Inventory (Dowd Therapeutic Reactance Scale; 12/13/2013).

Beck Depression Inventory (BDI; 12/13/2013).

Beck Hopelessness Scale (BHS; 12/13/2013).

Minnesota Multiphasic Personality Inventory—2 (MMPI-2; 12/21/2013).

Sarason Social Support Questionnaire (SSQ; 12/21/2013).

BEHAVIORAL OBSERVATIONS

Throughout the evaluation, A. G. was articulate, introspective, and cooperative. While acknowledging discomfort in talking about her sexual orientation

and religious feelings, she was quite forthcoming when questioned directly. A valid MMPI-2 profile, along with her willingness to cooperate and introspect, suggests that the current evaluation presents a valid picture of her current level of functioning.

RELEVANT HISTORY

Personal/Social. A. G. was raised in a middle-class, Jewish family. She was the older of two children, having a brother who is one and a half years her junior. The family always maintained at least a superficial religious identity and a facade of happiness. However, she reported that, behind this facade, there were significant underlying family conflicts. Religion has always been a source of conflict for her as has been her sexual orientation. Moreover, she reported a long family history of mental illness and interpersonal conflict. While she described her parents as emotionally stable, both her parents' families have histories of psychiatric disorders. Her mother was the oldest of nine children; an uncle died in a halfway house with a diagnosis of schizophrenia, another was diagnosed as having bipolar disorder, and still another was diagnosed as having depression. On her father's side, at least one uncle is reported to have had a major depression that was treated with antidepressants.

In her own personal history, A. G. reported that she always felt confused about her sexual orientation. At age 17, she received a proposal of marriage. She declined but he persisted, and she went to live with an uncle to escape his advances. He pursued her and finally, against her "better judgment," he talked her into marriage. The newlyweds moved to Metroville to be with her family, but problems persisted and they separated after about a year. By that time, she had given birth to a daughter. She moved in with her parents, but long-standing conflicts with her mother became more frequent. When her husband contested and prevented the culmination of the divorce, A. G. moved out of the family home and went to work, leaving her daughter with her mother. A. G. blamed her parents for her failed marriage and refused further contact. She did not see or speak with her daughter for two years.

After a period of estrangement from parents and daughter, the patient was contacted by her attorney, who informed her that A. G.'s mother could no longer raise the baby. Her husband was also informed, and he demanded that A. G. reconcile with him to raise the child. A. G. agreed to do so if they would move to Betterville to make a fresh start. Shortly after moving, A. G.'s husband became disillusioned and returned to Metroville, leaving her to raise the child. It was very shortly thereafter that she acknowledged to herself that she was a lesbian. She subsequently engaged in a series of brief lesbian affairs and adopted a "secret life" in which she prevented her parents and husband from an awareness of this emerging sexual orientation.

Still being unable to raise her daughter and work, the client gave up the child to a foster family for temporary care. After a few years, she initiated an

effort to again assume care of the child. Concerned about raising the child in a lesbian relationship, she accepted the proposal of marriage from a man who knew about her lesbian lifestyle. He nonetheless agreed to adopt the child and allowed her to continue her lesbian affairs. Their marriage lasted 23 years and produced two sons. Although unsatisfied with her dual life, she waited until her younger son graduated from high school before she left the marriage and began to pursue lesbian relationships exclusively and openly. She met her current lover in 2006. This relationship continues to be close although they ceased sexual contact approximately six years ago.

HISTORY OF PRESENTING PROBLEM

A. G. reported symptoms related to agoraphobia, panic attacks, and dissociation. She has a long history of panic attacks without agoraphobia, dating to age 12. The first panic episode occurred when she was babysitting for a family friend. She suddenly hyperventilated, began experiencing heart palpitations, and became afraid that she was going to die. She ran into the street yelling for help, but no one heard her or tried to assist her. The situation was resolved by exerting “self-control.”

After her initial panic attacks began, they gradually increased to a frequency of about once per week throughout her teenage years. To protect herself from feared panic and what she perceived as possible death, she frequently slept with her parents and confined herself to known places and locations. At their worst, her panic attacks involved physical symptoms such as nausea, shortness of breath (hyperventilation), and dizziness as well as cognitive symptoms such as fears of losing her mind, dying, of being overwhelmed, and unspecified danger. However, she learned to control these symptoms over time by avoiding such activities as going out, driving, and socializing with groups. These efforts have been successful in that A. G. reported that she had been asymptomatic for agoraphobia and panic for 31 years.

She currently reports that she has become apprehensive about travel and social activity, but the symptoms are confined to initial anticipatory anxiety, gastrointestinal distress, and headaches but with no heart palpitations, shortness of breath, or fainting. She prevents more extensive symptoms by avoiding travel and through a variety of distraction procedures. When she begins to experience the onset of panic, she calls someone or begins to read an interesting book. Her contacts with other people at these times do not include a disclosure of or discussions about the panic but are reported to be simply methods to involve herself with others and to take her mind off her feelings.

She reported that the current symptoms are mild in intensity and include a general discomfort with traveling, difficulty feeling comfortable when alone, and a general heightened sense of vulnerability and apprehension until she is able to return home. She continues to avoid night travel and avoids being alone, if possible, to prevent the associated anxiety.

Since 1996, several dissociative episodes have occurred, which she believes were precipitated by her decision to openly acknowledge her homosexual orientation. The first instance followed a sexual encounter with her current partner during a vacation. After the sexual act, the client experienced an apparent fugue state. She became disoriented, was unable to recall personal information such as that her parents were deceased, and engaged in distraught communication with her lover about "Why am I here." The episodes have subsequently recurred several times: They come on suddenly and without warning and she subsequently has no memory for the events. They uniformly follow a lesbian sexual encounter, and if her partner remains with her during this period (sometimes up to several hours), the fears gradually subside. However, after these dissociative states, she reported having a sense of helplessness, hopelessness, confusion, headaches, and nausea that sometimes lasted for several days. She has been able to successfully avoid these episodes by not engaging in sexual activities for nearly six years.

History of Treatment. A. G. was first treated and hospitalized in 1983 because of agoraphobia. There have been no subsequent hospitalizations. However, she has entered into two treatment relationships in the years since. Her current medication is managed by a psychiatrist, who is treating her with Xanax, Tagamet, and Paxil. She reported that since being on the medication, she has been excessively sleepy and has a difficult time staying awake during the day. She also has experienced an increase in stomach difficulties and diarrhea. She was also treated for a short time in 2003 by an internist and psychiatrist. At that time, she was given tricyclic antidepressants. These drugs produced hallucinations and were discontinued shortly after initiation.

A review of this woman's symptom history also reveals that she has had substantial periods of time in which she has been asymptomatic for fugue states, panic attacks, and agoraphobia. She reported that between August 2011 and June 2012, she was the "best ever." She was able to travel alone, found life enjoyable, and experienced no episodes of discomfort or fear. More recently, she has gradually become more depressed and dysphoric as well as fearful, although there was no obvious precipitator for these feelings.

Medical History. A. G.'s medical history is unremarkable. She currently takes Tagamet for stomach distress and Xanax for anxiety. Aside from some loss of hearing and psychophysiological symptoms, she acknowledges no significant medical problems.

INTERPRETATIONS AND IMPRESSIONS

Intellectual Ability. While a formal assessment of intellectual level was not undertaken, both A. G.'s verbal conceptual skills and oral presentation suggest at least average and probably bright-normal intellectual performance. Her ideation is dominated by preoccupation with ways to avoid uncomfortable feelings along with concerns with physical symptoms. Collectively, this results in

mild impairment to her cognitive efficiency. Her verbal processes are organized, circumstantial, and occasionally dominated by topics about which she has pressing concerns; but they reflect no disorganization, no memory impairment, and moderately well-developed associative and abstract reasoning processes. While she is oriented in all three spheres and manifests no significant mental impairment, she notes having always been concerned with the potential loss of mental functions.

Personality and Symptom Patterns. A. G. experiences ambivalent personality organization, with moderate disturbances to her functional adaptation. Her dominant conflicts involve strong needs for dependency counterbalanced by equally strong strivings for self-definition. A. G. denies dysphoria, depression, and anxiety. She complains of poor sleep, loss of energy, and lack of motivation. Formal assessment confirms the presence of vegetative signs (increased appetite, variable sleep, social withdrawal, loss of interest, reduced libido), consistent with the presence of mild to moderate depression without subjective dysphoria. Trait anxiety levels are within the normal range for her age, and subjective depression is only mild, with the dominant symptoms being psychophysiological. Her affect is appropriate though somewhat variable. Affective responsivity is both dysthymic and blunted.

Coping Style. A. G.'s mood disturbance reflects a chronic condition, against which she has constructed a variety of rigid and brittle defenses. She is excessively sensitive to environmental signals of threat and, at the least suggestion of emotional arousal, engages in both direct and cognitive avoidance patterns. The result is that she prevents the intensification or even emergence of feelings that might overwhelm her. While protecting her somewhat from subjective sensations of anxiety and dysphoria, A. G.'s defenses are not sufficiently strong to prevent the emergence of a variety of secondary symptoms. Denial, phobic avoidance in the face of anticipatory cues, self-criticism, compartmentalization, and somatization are among her most frequently used defenses. As threat intensifies, her fragile denial deteriorates, and both somatization and direct avoidance predominate. Thus, acute stress evokes a variety of stress-related somatic symptoms and phobic behaviors that provide expression for her denial of anxiety and depression.

Her coping style involves both passive and active efforts to reconcile these strong drives. Thus, while she seeks approval and confirmation from others, even to the point of excessive subservience in which she gives up personal strivings, this is frequently a futile effort to ensure the presence of other people in her life. Indeed, these efforts are usually designed to compensate for a host of covert rebellious and angry impulses and by overt efforts to be autonomous and self-guided. Thus, efforts to achieve self-fulfillment and autonomy are followed by guilt, self-doubt, and shame in which fear and withdrawal dominate. These latter symptoms, however, may be so demanding of attention that they are the functional equivalent of interpersonal anger, hurt, and resentment. Thus, her pattern of phobic anxiety and dissociation has led to sexual withdrawal and physical dependency. Her withdrawal may represent an indirect expression

of anger yet also be a compromise between asocial impulses and needs for approval. Unfortunately, this compromise also includes low self-regard and restricted mobility. Another consequence of this pattern is the current low level of available others to provide support. In spite of this, A. G.'s satisfaction with the level of interpersonal support available from her significant other is good and suggests the availability of this individual as a support in any treatment program.

Client Strengths. A. G. has good awareness of many of the dynamics underlying her condition combined with a moderate amount of ego strengths. Her cognitive abilities are estimated to be in the average to high-average range, and her thought processes are focused and intact. Despite being quite avoidant, her level of distress is at an optimal level to assure her motivation to become engaged in therapy. Her relationship with her partner is both long term and quite supportive.

DIAGNOSTIC IMPRESSION

The diagnosis of anxiety disorder, specifically agoraphobia (300.22), is based on history rather than current symptomatology. A major differential question which should be ruled out is the possible presence of major depressive disorder (296.3).

SUMMARY AND RECOMMENDATIONS

A. G. is a 70-year-old, divorced female with 13 years of education with symptoms related to agoraphobia, panic attacks, and dissociation.

1. Nature and dynamics of the problem. A. G.'s symptoms are of a complex and long-standing nature founded more in dynamic and early developing interpersonal expectations and conflicts than in symptom-contingent events. These core conflicts seem to be largely founded in postpubescent strivings to resolve needs for autonomy and dependency. A. G.'s level of functional impairment is moderate. Numerous areas of functioning are affected, and this, coupled with the chronicity of the condition, suggests the need for long-term treatment. A. G.'s level of distress is well contained since it falls in the average or even below-average range compared with other patients who seek treatment. While her distress increases significantly when exposed to immediate threat, she quickly compensates and so is well versed at avoidance.
2. Client strengths. Relevant strengths of A. G. are that she has good awareness, intellectual level that is both intact and in the average to high-average range, experiences an optimal level of motivational distress, and is in a quite supportive long-term relationship.

3. Treatment plan. The dynamic nature of the associated conflicts, and their role in maintaining systemic dysfunction in her relationships with significant others, suggests the need to combine a symptom-focused treatment with efforts to resolve fundamental conflicts. The initial focus of treatment should be on reducing territorial apprehension, with a concomitant increase in social involvement and independent functioning. After initial symptomatic improvement, further interventions should focus on A. G.'s pattern of rebelliousness, which seems to be intertwined with self-incrimination, guilt, and withdrawal. In particular, these interventions might emphasize confirming needs for both autonomy and acceptance, along with greater insight into this pattern.

The patient's level of defense and personal control is sufficient, and the level of subjective despair and hopelessness is within a range that suggests that outpatient care is appropriate. There is no evidence of direct risk to self or others. Anxiolytic or antidepressant medications are contraindicated because of her relative degree of control over her symptoms, combined with the high potential for somatic side effects. Individual treatment may allow a more selective and intensive focus on problematic behaviors. Individual therapy would also be likely to prevent the operation of direct avoidance of discomfort when compared with group treatment.

Despite an optimal level of motivational distress, she may experience difficulty sustaining sufficient motivation for therapy due to her strong avoidant patterns.

Thus, interventions that confront or expose her to feared and avoided circumstances may be helpful to desensitize her to anxiety as well as to maintain her level of motivation to continue treatment.

A. G.'s coping style vacillates between being primarily impulsive and externalizing, to being self-critical and internalizing. This pattern of cyclic coping suggests the need to address her problems at both a behavioral and an insight level. When her impulses and direct avoidance dominate, behavioral strategies should be emphasized. During phases in which she is more introspective and self-blaming, insight-oriented interventions are likely to be more effective. Given the unsustaining nature of her subjective distress, abreactive and sensate-focused, cathartic interventions may prove to be especially helpful during these more introspective phases.

A. G. manifests a pattern of superficial compliance and more covert resistance to the directives of help givers. Thus, special attention should be given to developing a trusting relationship. Even if this is achieved, however, she would still be expected to undermine direct suggestions and specific assignments. The most effective approaches, then, would be collaborative interventions emphasizing clear behavioral change, contingency contracting, or paradoxical interventions such as symptom prescription and "no-change" directives. Particular attention may be given to predicting the exacerbation of physical and phobic

symptoms following intense sessions because these sessions may mobilize her resistant impulses in an asymptomatic direction.

Collectively, the symptomatic aspects of the patient's fears and phobias may be susceptible to a combination of structured exposure procedures, cognitive restructuring, and interoceptive awareness (Craske & Barlow, 1993). These procedures circumvent patient resistance by virtue of their reliance on self-monitoring as well as being both symptom and behaviorally focused. The more thematic and dynamic aspects of A. G.'s problem may be addressed by initiating work that specifically mobilizes her anxiety in motivational directions. Confrontation with feared material, along with the use of procedures such as two-chair work and imaginal reliving of unsettling relationships, may be helpful. Imaginal confrontation might be initiated with images and memories of disapproving parents, children, and other significant others, the goals of which may be to help her tolerate discomfort and disapproval. The procedures outlined by Daldrup, Beutler, Engle, and Greenberg (1988, *Focused Expressive Psychotherapy*) for working with the overcontrolled patient may also be particularly helpful.

Source note: Report written and submitted by Larry Beutler, PhD, ABPP, Palo Alto University, Palo Alto, CA.

Legal Context

NAME: Joe Competent
AGE (Date of Birth): 49 (1/10/65)
SEX: M
ETHNICITY: Hispanic
DATE OF REPORT: 6/25/2014
NAME OF EXAMINER: Frank Clinician, Ph.D., ABPP
REFERRED BY: John McReferral
Legal Context

REFERRAL QUESTION

Mr. Competent is a 49-year-old, right-handed, divorced, Hispanic male who has been previously diagnosed with paranoid schizophrenia and “learning disabilities.” I understand that, following apprehension for a suspected burglary, he waived his Miranda rights, was interrogated by the police, and confessed to having broken into a house. I understand that you would like me to determine his cognitive capacity and address the following questions:

1. What is the extent he could comprehend the Miranda rights that were read to him?

2. How competently would he have knowingly and intelligently participated in his interrogation?

EVALUATION PROCEDURES

Clinical interview, Neuropsychological Symptom Checklist, Wechsler Adult Intelligence Scale—IV, Wechsler Memory Scale—IV, Wide Range Achievement Test—IV (WRAT—IV), Controlled Oral Word Association Test, Bender, Bender—2 Memory, Minnesota Multiphasic Personality Inventory—2 (MMPI-2; questions were read to him), Rey 15 Item, Test of Memory Malin-gering (TOMM), transcripts of police interrogation, competency evaluation by Jonathan Smith, M.D. (7/22/2012), competency evaluation by Patricia Jones, M.D. (6/15/2012). Total face-to-face evaluation time 5 hours, 40 minutes.

BEHAVIORAL OBSERVATIONS

Mr. Competent was seen in the Metrotown County Jail. He was dressed in prison uniform but without restraints. He was missing most of his teeth and the left side of his mouth was slightly lower than the right side. His ability to pronounce words was somewhat poor due to what he described as a “hairlip.” His affect was somewhat flat but was otherwise normal. His thoughts were sometimes disorganized and his responses were fairly tangential. However, there did not seem to be any obvious signs of delusions and he denied having any hallucinations. Despite this, he did appear somewhat suspicious and on two occasions asked who would be reading the report I was going to prepare about him. He seemed to have fairly poor insight into his psychiatric difficulties and felt that his suspiciousness toward others was simply due to his having been assaulted when he was in his 20s. When asked why he was taking medication and the impact it had on him, he stated that it made him tired but gave no indication it had any other impact on his emotional or cognitive functioning. The responses he provided on his personal background were quite vague, and on several occasions he altered some of the details. For example, when asked if he was married, he at first said no and then said he had been married but it was when he was in jail. He then said he had gotten divorced but soon afterward said “no, I never did get divorced.” Due to these discrepancies, I think he was a fairly poor historian. He did seem to understand the questions I asked him and was usually able to follow instructions. However, on many occasions he needed to have the instructions repeated. On one occasion he appeared to not comprehend the instructions to a task despite repeated attempts at clarification. He did seem to give his best efforts toward the tasks presented to him.

Given the above behavioral observations, I believe some of the history he provided may have been somewhat inaccurate due to his disorganized and tangential thoughts. As a result, I relied on a review of records to develop a complete history. He scored low on measures evaluating the validity of his

responses to cognitive testing (so-called fake bad tests). For many clients this might indicate an exaggeration of deficits. However, in Mr. Competent's case, these low scores were consistent with his history, previous evaluations, and quite low performance on other cognitive tests. As a result, I feel the results of testing were generally an accurate assessment of his current level of functioning. However, given his ethnic background, some of the results may be a slight underestimate of his potential. As a result, this has been taken into account during the interpretation of test scores.

BACKGROUND

(Note: Background history has been significantly abbreviated and altered to ensure anonymity.)

PERSONAL/SOCIAL: Mr. Competent stated that he was born and grew up in Anytown. He stated that he had two brothers and six sisters, but he was unclear as to whether these were biological or foster siblings. English was the main language spoken in the home. Medical records did not refer to him having lived with a foster family. They also indicate he was born in Anytown but about the age of 9 he and his family moved to Metrotown. Mr. Competent reported that his "foster mother" died at age 79 and he doesn't know if his "foster father" is still alive since he said he last saw him in 1993. It is quite possible that when he refers to his "foster" parents, he is actually referring to his biological parents. Mr. Competent said that his father worked on a ranch and his mother cleaned houses.

During high school Mr. Competent stated that he "got along with people" and had many friends. He further stated he liked PE classes and going to parties. Forensic and psychiatric records indicated he had a number of difficulties with the law and spent several years in a Forensic Youth Authority facility beginning in 1983. Medical records indicated he had been married in 1987, but he stated that he did not have any children. In contrast to this, his medical records indicate that he did have three children. Medical records also indicate he has no contacts with either his wife/children or with his family of origin.

ACADEMIC/VOCATIONAL: Mr. Competent stated that during high school he went to "special classes" since he had a difficult time with reading. He said that his most difficult subjects were English, History, and Math. He stated that his grades were a combination of As, Bs, and Cs. He described himself as a "slow learner" and mentioned that he had a "learning disability." Some records indicated he went to Metrotown High School whereas other records indicated he attended Northbridge Special Education School. It should be noted that for at least part of this time he would have attended school at a Forensic Youth Authority facility. No educational records, including formal psychological assessments, were available for review. Employment has been intermittent, rarely lasting for more than a few weeks. He clarified he "didn't quit, I just didn't go back to work."

MEDICAL: He denied ever having had a tumor, stroke, unusually high fevers, exposure to neurotoxic substances, and described his use of drugs or alcohol as being moderate (with no use of “hardcore” drugs). He has evidently smoked one pack of cigarettes a day for much of his life.

PSYCHIATRIC/LEGAL: Early psychiatric history prior to 1993 was unavailable. It is thus not clear how his first symptoms began and how they were expressed. Review of forensic and psychiatric records indicated he has usually carried a diagnosis of paranoid schizophrenia. Additional diagnoses have included schizoaffective and bipolar disorders. Medication has included Haldol, Zyprexa, Prolixin, Depakote, Risperdal, Cogentin, and Paxil. Complications with his treatment have been poor awareness that he has a mental illness, poor comprehension of his commitment and discharge criteria, difficulty taking care of his health needs, poor medication compliance, social isolation, and suspiciousness. At times he has been able to state that he has a mental illness and that his medication helps him. Most of the time he has expressed that he does not feel he has a mental illness. Substance abuse has been identified as a difficulty, but there was unclear documentation that this had actually been a problem for him. During my interview with him, he minimized any substance abuse, but it is unclear whether this was due to poor insight or whether he simply has not had a substance abuse problem. Treatment has been partially but intermittently successful. One area of success has been that he only rarely seems to experience full-blown psychiatric symptoms such as delusions and hallucinations. Mr. Competent stated that current medication includes Risperdal (4 mg/1Xday) and Cogentin (2 mg/1Xday). Past mental health records are all consistent with him having quite low verbal comprehension, poor insight, low academic achievement, and generally low intellectual abilities. However, these were based on an impressionistic analysis rather than on more formal, precise, psychological measures.

SUMMARY OF ALLEGED OFFENSE

On the evening of 5/25/14, Mr. Competent was apprehended for “prowling.” He waived his Miranda rights and, upon questioning, confessed he had been in a house that had been broken into. He was then transferred to Metrotown County Jail, where he is currently incarcerated. He now feels he was manipulated by the police into confessing that he had made an attempted burglary.

INTERPRETATION AND IMPRESSIONS

General Level of Intellectual Function. Overall level of functioning was in the extremely low range or the lower 1% of the population when compared with his age-related peers (Full Scale IQ = 62). This is in the mildly mentally retarded range. Both his verbal knowledge/reasoning abilities as well as his ability to

process visual spatial information were equally low. Given his history and previous evaluations, his low intellectual level is likely to be of a long duration.

Attention and Concentration. A noteworthy difficulty was that he has an extremely difficult time paying attention. This means that he is likely to not focus on conversations and to have extreme difficulty trying to learn new material. His formal score in this area was in the 0.1 percentile, which means that only 1 person in 1,000 would perform in this low a range.

Memory. Overall memory was in the extremely low range or the lower 0.1 percent of the population (Immediate Memory Index = 55). This means that he would have a quite difficult time remembering things that people have told him or events that had occurred. For example, he was only able to repeat a maximum of three numbers that had been read to him. His ability to recall a short paragraph that was read to him was extremely limited.

Verbal Skills. Verbal abilities were in the lower 2% of the population (Verbal Comprehension Index = 68). His word knowledge was quite limited as was his fund of information. For example, he defined “today” as “in the past” and thought there were “6” days in a week. His ability to use everyday reasoning to understand common situations was also quite low. Even though he could name common objects, such as a cross and a square, he could not spell them.

Achievement Level. Formal assessment of his reading indicated he could read up to the fourth-grade level. His spelling was at the first-grade level. Although he was able to spell “cat” and “hut,” he was unable to spell “leg” and “shirt.” Arithmetic was at the third-grade equivalent. For example, he was able to add and subtract numbers but was unable to do simple division (he thought 20 divided by 4 equaled 30). These scores suggest that he is functionally illiterate in that he would be unable to read newspapers, magazines, books, legal documents, or financial information.

Nonverbal Skills. Mr. Competent’s nonverbal abilities were relatively better than his memory and attention. This means that a relative strength is his ability to understand and problem-solve information he sees. However, his nonverbal abilities were still in the lower 2% of the population (Perceptual Reasoning Index = 72). In addition, his ability to reproduce a series of simple designs was clearly in the impaired range.

Speed of Information Processing. The speed by which Mr. Competent processes information is in the lower 1% of the population. This means that explanations and questions would need to be given with extreme care. In order for him to comprehend and solve problems, he would also need to be given extra time. It should be noted, however, that some of his psychiatric medication is sedating so this may have exaggerated Mr. Competent’s low processing speed. To compensate for this, I tried to give him tasks requiring speed, concentration, and memory during times in which the impact of his medication would be minimal.

Psychiatric. Review of personality and symptoms indicated an unusual pattern of scores on the MMPI-2. He appeared to exaggerate his symptom complaints. At the same time he minimized them by emphasizing his positive

qualities in an unrealistic manner (e.g., he denied ever getting angry and stated that he never felt like swearing). I think this may be due to not adequately understanding the questions since they are at a fifth-grade reading level. At the same time, previous evaluations seem to indicate a style in which he both minimizes his symptoms and yet he still expresses fairly severe psychiatric illness. Despite this, his overall profile was still consistent with his history. There were moderate elevations indicating disorganized thinking, unusual thought processes, tangential thinking, periods of high energy and euphoria, and he is also oversensitive, mistrustful, and suspicious. However, at the time of assessment, he was cooperative and did not seem to have any overt delusions or hallucinations.

SUMMARY AND OPINION REGARDING QUESTIONING

Mr. Competent is a 49-year-old, right-handed, married male with a history of psychiatric illness and legal difficulties. Past treatment has been complicated by low insight, low intellectual level, social isolation, intermittent assaults, suspiciousness, difficulty caring for his medical conditions, and poor compliance. My evaluation found him to be functioning in the mildly mentally retarded range (overall IQ = 62, or the lower 1% of the population) with specific difficulties in verbal comprehension, verbal reasoning, impaired attention, poor memory, low speed of processing information, and impaired ability to construct designs. Academic achievement ranged from a first-grade equivalent in spelling to a fourth-grade equivalent in reading. This is sufficiently low for him to be functionally illiterate. He would also have a difficult time understanding concepts that had even a mild level of complexity. This means that he would need to have information presented in an extremely simple form with clear explanations of the implications of his responses. Due to his poor attention and slow speed of processing information, considerable extra time would be needed for him to comprehend and respond to information. However, he is able to understand things generally when information is explained in simple, uncomplicated terms. Psychiatric symptoms suggest disorganized thinking, unusual thought processes, tangential thinking, oversensitivity, suspiciousness, and periods in which he experiences high energy and euphoria.

1. Concerning waiver of Miranda warnings and competency of waiver:
Examination of Mr. Competent's reading and language skills in comparison to the comprehension and reading levels required to fully understand the Miranda warnings indicate a significant discrepancy. His verbal abilities and overall mental abilities are in the mild mentally retarded range (lower 1% of the population) and range between the first- and fourth-grade levels. In contrast, grade level for the Miranda warnings typically fall between the sixth- and eighth-grade level (see research review by Helms, 2003). Based on these considerations, it is

my opinion, expressed within a reasonable degree of psychological certainty, that Mr. Competent's waiver of his Miranda rights to the Metrotown police was unlikely to be "knowing and intelligent." I believe it was unlikely that he was able to waive these rights with a full awareness of the rights being waived and the consequences of the decision to wave them.

2. Concerning his ability to have knowingly and intelligently been able to respond to interrogation:

Mr. Competent's low intellectual level along with his chronic mental illness would also have made it difficult for him to knowingly, intelligently, and voluntarily participated in his interrogation. He would have been unlikely to have withstood the stress of interrogation and also to understand the meaning and implications of seemingly "friendly" police. In other words, he would have a difficult time distinguishing between the fact and the appearance of friendliness. Research indicates that a high proportion of persons who are mentally retarded respond with "yes" answers regardless of the content of the questions (see Sigelman et al., 1981; "When in Doubt Say Yes: Acquiescence in Interviews with Mentally Retarded Persons"). Research also indicates that both mentally ill and mentally deficient persons are those who are most likely to provide what are often referred to as false confessions (Clare & Gudjonsson, 1995). They are much more susceptible to leading questions, confabulate more, and are more acquiescent to interrogators (Clare & Gudjonsson, 1995). They are also more likely to believe that falsely confessing will have little or no consequences. Thus Mr. Competent has an ideal set of characteristics that would make him at high risk of providing such "false confession."

Note that the opinions and conclusions herein are clinical in nature, and they do not represent legal conclusions. Ultimate legal questions are solely for the court to decide. I appreciate the opportunity to have been of service. Should you have any questions about the above client, please do not hesitate to contact me at your convenience.

Gary Groth-Marnat, Ph.D., ABPP, ABAP
California State License (PC XXXX)

REFERENCES

- Clare, I. C. H., & Gudjonsson, G. H. (1995). The vulnerability of suspects with intellectual disabilities during police interviews: A review and experimental study of decision-making. *Mental Handicap Research*, 8, 110–128.

Helms, J. L. (2003). Analysis of *Miranda* reading levels across jurisdictions: Implications for evaluating waiver competency. *Journal of Forensic Psychology Practice*, 3, 25–37.

Sigelman, B. E., Spanhel, C., & Schoenrock, C. (1981). When in doubt say yes: Acquiescence in interviews with mentally retarded persons. *Mental Retardation*, April, 53–58.

The Educational Setting

NAME: David W.

DATE OF BIRTH: XX/XX/XX

GENDER: Male

DATES OF ASSESSMENT: February 4, 12, 16, 2015

DATE OF REPORT: February 25, 2015

EXAMINER: A. Jordan Wright, Ph.D., ABAP

REFERRAL QUESTION AND PRESENTING PROBLEM

The client is a 12-year-old, right-handed, Caucasian boy who is in the sixth grade at XXXXXX, a private school in XXXXXX. He was referred for an assessment primarily to evaluate what his learning needs are and likely will be in the future. His parents reported that he has “executive functioning challenges.” Specifically, they noted a “disconnect” between his ability in school and his “engagement and execution.” They reported that when he is interested in a subject, he is an A or B student; however, more frequently, he is not engaged in his academic work and struggles with his grades. Most notably he is significantly struggling with math. His parents noted that he has difficulty with testing, seemingly related to time, and with following directions fully. His problems following directions extend beyond tests, as he has difficulty following complex instructions in class. Additionally, he reportedly reads about things in which he is interested (especially space and space travel), but he does not tend to read otherwise. He is currently receiving a great deal of academic support, including from his current teachers, his parents, and a tutor who was a teacher from his previous school.

The client and his parents reported that he had some behavioral and social problems earlier in the school year, related to some fighting with peers, but these have reportedly disappeared with some intervention (from school, parents, and his current therapist). The purpose of the evaluation is to clarify the nature of the client’s difficulties and to make recommendations for improving his functioning, especially at school.

EVALUATION PROCEDURES/SOURCES OF INFORMATION

Review of previous assessment (from 2011)
 Clinical Interview with client
 Collateral Interview with client's mother and father
 Bender Visual-Motor Gestalt Test, Second Edition (Bender-2)
 Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V)
 Conners' Continuous Performance Test, Second Edition (CPT-2)
 Repeatable Battery for the Assessment of Neuropsychological Status
 Update (RBANS)
 Wisconsin Card Sorting Test: Computer Version 4 (WCST-IV)
 Trail Making Test (Trails A and Trails B)
 Wechsler Individual Achievement Test, Third Edition (WIAT-III)
 Rorschach Performance Assessment System (R-PAS)
 Behavior Assessment System for Children, Second Edition (BASC-2)
 Parent Report Scale—Adolescent (PRS-A)
 Teacher Report Scale—Adolescent (TRS-A)
 Self-Report Scale—Adolescent (SRS-A)

BACKGROUND INFORMATION

History of Presenting Problem. The client's parents reported that he was slightly slow to read when he was first learning, seemingly because of a lack of interest in it. They stated that academic development was described as "step-wise" rather than linear, in that skills seemed to develop suddenly and rapidly at different times. Notably, his parents reported that he has tended to have an excellent vocabulary. They reported that the client really began struggling with school when homework became "real" and focused on developing actual academic skills. As the work in school got harder, he reportedly "didn't grow with it." In 2011, he was evaluated by XXXXXX, Ph.D. (dated 4/23/2011) for cognitive and neuropsychological functioning, and the report found generally intact cognitive functioning, with some minor difficulties in planning, response inhibition, and sustained attention, though none of these areas was grossly impaired. He switched schools after the fifth grade (from XXXXX to XXXXX), and it was decided that he should repeat the fifth grade at his new school.

Other Relevant History. The client is the younger child of two and currently lives with his mother, father, and older sister, who is 14 years old. His mother works as a XXXXX, and his father works in XXXXX. The client's parents reported that he has many hobbies, including learning about space and space travel (about which he is passionate) and playing tennis, soccer, and other sports. He reportedly enjoys building things (with Legos) and playing with design software as well.

According to his parents, there were no known difficulties with the client during pregnancy or birth/delivery, and he reportedly met all developmental milestones adequately, though he was slightly late but fast to walk and talk.

He reportedly potty trained himself. He has no past or current major medical problems and no significant history of neurological problems. His father had cancer about 10 years ago, which is in remission, and his parents denied any other major medical problems in the immediate family. They reported that his maternal grandfather heavily used drugs and alcohol and that his maternal grandmother was “maybe depressive,” but they denied any other psychiatric problems in the family. The client is not currently on any medications.

The client is reportedly socially healthy, with significant friendships. Although he had some social behavioral difficulties earlier in the present school year, these have subsided. He reported enjoying spending time with friends, especially on the soccer team at school. He currently receives cognitive-behavioral therapy at XXXXXXXXXXXX, focused on the behavioral problems he exhibited earlier in the school year.

BEHAVIORAL OBSERVATIONS/MENTAL STATUS EVALUATION

The client was appropriately dressed and groomed throughout the assessment, and he looks his stated age. He showed up about 10 minutes early to each appointment, accompanied by his father. He was somewhat slow to warm up with the assessor, talking little and yawning throughout the beginning of the first session. He became more talkative and friendly as the assessment progressed, spontaneously talking and making better eye contact. He seemed to persist and give full effort on all activities, except at times when he felt a test item would take too much time or effort and gave up. His eye contact was adequate, and he exhibited no psychomotor agitation or retardation. His speech and language were within normal limits. His mood was reportedly “good,” and his affect was consistently mood-congruent. His thought process was generally clear and goal-directed. His thought content was seemingly free of delusions, and he did not report anxious or depressive thoughts. The client denied hallucinations and suicidal and aggressive/homicidal ideation. His attention and concentration were adequate throughout, and his memory functioning appeared intact. His insight was adequate, though he reported feeling that some people find him “annoying,” but he does not know why. His judgment seemed generally good in the moment. Based on his seemingly effortful attempts on the tests administered, the current assessment is likely a good estimate of his current functioning.

OVERALL INTERPRETATION OF TEST FINDINGS

Cognitive Functioning

The client’s overall ability was in the average range compared with others his own age (WISC-V FSIQ, 63rd percentile, better than 63% of his same-age peers). However, there were some notable variations in his abilities.

Strengths

Verbal Ability. His verbal ability is high average compared to others his age (WISC-V Verbal Comprehension Index, better than 86% of his same-age peers), with specific strengths in his ability to express himself clearly and to think abstractly using language in a complex and sophisticated way (WISC-V Vocabulary, better than 91% of his same-age peers; WISC-V Similarities, better than 75% of his same-age peers). He can understand and use language to express himself in an effortless and sophisticated way.

Nonverbal Ability. He showed no difficulties in his fine motor control (Bender-2 Motor Subtest, better than 76–100% of his same-age peers), so he should have no difficulties with the neatness of his handwriting or precision of drawing. His visual-perceptual abilities (WISC-V Visual Spatial Index, better than 63% of his same-age peers) and visual-motor integration (Bender-2 Copy, better than 61% of his same-age peers; WISC-V Block Design, better than 50% of his same-age peers) were also adequate. As such, he perceives visual information well, and can copy designs and pictures well.

Learning and Problem Solving. Further, he exhibited no difficulties in his ability to learn new information, hold it in short-term memory, concentrate, and manipulate that information to produce some result or reasoning outcome (WISC-V Working Memory Index, better than 50% of his same-age peers), as well as his ability to solve novel problems that use very little prior knowledge or expertise (WISC-V Fluid Reasoning Index, better than 50% of his same-age peers) and his ability to solve problems, change strategies, and inhibit his impulses (also known as executive functions; Trails B, better than 61% of his same-age peers; WCST-IV Total Errors, better than 50% of his same-age peers). He has no difficulties solving problems with new information or changing strategies toward problems if his chosen strategy is not working.

Weaknesses

Attention. The client exhibited fluctuations in his abilities to sustain his attention and to work quickly and accurately, performing well at some points and poorly at others, even though the tasks were either similar or the same. On a measure that was given to him twice that required him to sustain his attention and persist during a boring task, during the first administration he had significant difficulty being persistent and consistent over time (CPT-2 Hit SE Block Change, better than only 5% of his same-age peers); during this administration he was yawning and seemed somewhat distracted. During the second administration of the exact same task, his persistence and consistency over time was better than 94% of his same-age peers; this administration was just after lunch, and he had better rapport with the assessor at this point.

Processing Speed. Similarly, on several tasks his speed of processing information and performing was significantly lower than others his age (Trails A, better than only 5% of his same-age peers; WISC-V Processing Speed Index, better than 18% of his same-age peers). However, his speed of processing information seems highly dependent on his motivation and present

state of mind, as on similar tasks he showed the ability to work much more quickly at one point than another (WISC-V Coding, better than 5% of his same-age peers; RBANS Coding, better than 61% of his same-age peers).

Memory. A final weakness in his cognitive abilities was his delayed (medium-term) memory. He exhibited no problems with any area of learning or immediately remembering information, including word lists, stories, and visual information (RBANS Immediate Memory, better than 70% of his same-age peers). His delayed memory (after about 30 minutes) on these same tasks was low average (RBANS Delayed Memory, better than 18% of his same-age peers; Bender-2 Recall, better than 14% of his same-age peers). His delayed recognition of earlier learned verbal information was adequate (RBANS List Recognition, better than 75% of his same-age peers). This shows that he learned the information, and on the easier task of recognizing what he learned, he had no difficulty. However, his ability to recall information after a delay is a weakness for him and should be the focus of extra attention when it comes to him studying and learning information for school.

Academic Achievement

In comparison to his average global cognitive potential and especially his excellent verbal ability, the client's performance within academic areas was consistently at or above average, which would be expected, except for his speed and fluency.

Strengths

Reading. He showed no difficulty in his reading ability (WIAT-III Total Reading, better than 82% of his same-grade peers), including good knowledge of phonics, word recognition, and comprehension (WIAT-III Pseudoword Decoding, Word Reading, and Comprehension, better than 61%, 86%, and 84% of his same-grade peers, respectively).

Writing. His writing ability was similarly high average (WIAT-III Written Expression, better than 84% of his same-grade peers).

Mathematics. His basic mathematical abilities were also high average compared to others at his grade level (WIAT-III Mathematics, better than 82% of his same-grade peers).

Weaknesses

Speed and Fluency. Where he struggled somewhat was his ease, speed, and accuracy of reading aloud (WIAT-III Reading Fluency, better than 34% of his same-grade peers) and his speed and accuracy of completing basic arithmetic (WIAT-III Math Fluency, better than 39% of his same-grade peers). His speed, ease, and accuracy of completing reading and math tasks are lower than would be expected, given his overall good abilities. However, given the fact that his ability to sustain attention and work quickly in general fluctuated significantly, these slightly lower levels of achievement make sense.

Emotional Functioning

The present assessment revealed that, in addition to the client struggling academically, he has difficulty tolerating frustration. The result of him having difficulty academically and not being able to tolerate the frustration of his school struggles includes some insecurity about himself. His low self-esteem, combined with his difficulty tolerating frustrating situations, has led to some acting-out behaviors, though these have subsided recently.

Academic Problems

Consistent with the cognitive testing findings from this assessment, the client is struggling academically at school. The R-PAS revealed that he is currently experiencing some form of stress related to his environment, and his TRS-A revealed that this stress is related to his learning problems. His SRS-A revealed that he is currently quite frustrated in school and dislikes it, feeling that his teachers do not understand him. His academic problems are taxing his ability to tolerate frustration and leading to a negative view of himself.

Low Frustration Tolerance

The client struggles to tolerate frustration, giving up easily and acting impulsively if and when he needs to persist on tasks. His R-PAS revealed that he approaches the world and problems in it by reacting spontaneously, rather than thinking through problems. His SRS-A revealed that he has a short attention span, but his PRS-A and TRS-A revealed that he is not easily distracted, suggesting that his short attention span is not the result of attentional problems. His TRS-A and PRS-A revealed that he is easily upset by frustrating situations, and his TRS-A revealed that he often acts without thinking. His SRS-A revealed that he gives up easily when frustrated, and his R-PAS revealed that his impulsiveness and giving up are the result of low general coping skills. This low frustration tolerance, combined with his academic difficulties, has led to a negative view of himself.

Low Self-Esteem

As a result of difficulty tolerating frustration and academic difficulties, the client has adopted a view of himself as inadequate and damaged. His SRS-A revealed a general sense of low self-esteem, as well as a strong sense that he is inadequate. His R-PAS revealed that his understanding of himself is problematic and characterized by feelings of vulnerability and being damaged in some way. His PRS-A revealed that at times he says that he hates himself, which is a manifestation of his low self-esteem. A result of his low self-esteem, combined with his poor frustration tolerance, is some acting out behavior.

Acting-Out Behaviors

As a result of low self-esteem and difficulty tolerating frustration, the client tends to act out behaviorally, though this has largely subsided with his current therapy. His PRS-A and TRS-A both suggested that he teases other adolescents, as well as breaks some rules and defies his teachers at times. However, these behaviors were reported earlier in the school year and are reportedly not currently happening at school or home.

DIAGNOSTIC IMPRESSIONS

Currently, the client meets criteria for a Nonverbal Learning Disability (NVLD), which is categorized in the *DSM-5* as an Other Specified Neurodevelopmental Disorder (315.8). Specifically, there are minor details that seem to go unnoticed (such as understanding every single piece of directions given to him), fluctuations in his ability to pay attention, academic difficulties (especially in math) that are easily mistaken for carelessness, attention problems that are predominantly limited to school, and some acting-out behaviors that are likely related to both stress and visual and tactile input. These are the major symptoms of NVLD. Notably, while attention is a problem, he does not exhibit a cognitive profile consistent with attention-deficit/hyperactivity disorder (ADHD).

Consistent with others who have NVLD, the client will be expected to struggle academically in the following ways:

- Difficulties with unexpected changes in routine or last-minute switches in scheduling.
- Difficulty generalizing previously learned information.
- Difficulty following multistep instructions.
- Slow processing speed.
- Becoming easily overwhelmed (especially because of slow speed of processing information), which can lead to shutting down or loss of motivation.
- Heightened sensitivity to sensory stimulation, with difficulty responding to multisensory information (such as simultaneous visual and auditory information).
- Often looking like they are not meeting their potential, because by all other measures they seem capable or better.
- Problems with longer-term memory of learned information.

SUMMARY AND RECOMMENDATIONS

The client is a 12-year-old, white boy who was referred for an evaluation to assess his learning needs and how they will likely best be met in the future. He is reportedly struggling with meeting his potential at school, especially in certain subjects like math. He has difficulty following complex directions and has problems on tests, especially timed tests. Although somewhat slow to warm up, he became friendly and engaged with the assessor during the process, and he seemed to give effortful attempts on all tests administered.

The client has a significant strength in his use of language and verbal ability, which has translated into above-average ability in reading, writing, and mathematics in general and should serve him well across different areas of life, especially in academic and professional settings.

In contrast to this strength, he showed fluctuation in his ability to sustain attention and his speed of processing information, revealing that these abilities are highly dependent on his present state of mind and motivation. Although he has the ability to sustain attention (concentrate) and process information quickly, at different moments in time there are thoughts, feelings, and attitudes that get in the way of these abilities. These difficulties have translated into difficulties with the speed, ease, and accuracy of reading aloud and the speed of performing basic arithmetic problems.

Additionally, he showed weakness in his ability to recall learned information after a delayed period of time. Along with his fluctuation in attention and speed of processing information, this difficulty will likely manifest in school and on homework, where he will likely display fluctuation in his performance in general and difficulty on exams.

His academic and cognitive difficulties make his already weak ability to tolerate frustration worse, and these together have led to insecurity about himself. This low self-esteem, unfortunately, leads him to give up on some of his academic work, rather than persisting, which only weakens his academic performance. Additionally, his low self-esteem, combined with his difficulty tolerating frustrating situations and his cognitive weaknesses, has led to some acting-out behaviors, though these have subsided recently.

Given the previous findings, the following recommendations are being made:

1. The *client's parents* should discuss the results of this evaluation with the staff of his school.
2. The client should be provided with the following *school and academic accommodations*:
 - He should be afforded extra time on all examinations, both standardized and within the context of school classes.
 - Assignments and exams with multistep instructions should be broken down to clear, step-by-step directions for successful completion.

Steps should be numbered and sequenced in the most efficient and logical path to successful completion possible. When possible, these steps should be presented only a few at a time.

- On assignments and exams, whenever possible, a teacher should check in with the client to ensure his understanding of all the directions.
 - Efforts should be made to minimize multiple sources of sensory stimulation; that is, efforts should be made for him to sit where there is as little background distraction as possible. Any method of decreasing visual stimulation during class should be attempted, as well as allowing him to fidget mindlessly with his hands as needed.
 - He should be allowed to take regular breaks throughout the day to decrease the occurrence of being overwhelmed and subsequently shutting down.
 - The teacher and client should together discuss and identify signs of overload so that they can observe them as they happen and intervene appropriately (e.g., take a break).
 - Homework assignments should be pared down to the least amount necessary to reinforce the learning from class lessons.
 - Teachers should adopt strategies to ensure that the client truly understands concepts (what he hears in class, what he has read) beyond being able to say it back or describe it at a surface level. Although he has excellent basic reading, surface reading comprehension, and even listening skills, remembering material for a longer period of time and especially applying learning to different situations will be more difficult for him. As such, even when it *seems* as if he understands, teachers may need to find ways to ensure that he has a deeper level of learning of material.
 - Teachers should be willing to offer additional verbal instructions, even when written instructions are clear, to ensure understanding.
 - The school should help identify a neutral, private space where the client can regroup and relax for short periods of time, as needed.
3. A *learning specialist* should help the client in the following ways:
- He should be given strategies and encouraged to practice these strategies to break down directions very carefully, rewrite them as necessary in a more broken-down, orderly way, and carefully follow through with them.
 - He should clearly articulate how old learning applies to new situations. For example, a learned mathematical concept should be thoroughly understood *and articulated* before it is applied to a novel type of problem.
 - He should practice articulating clearly what generalizations can be made from specific learning and situations.

- He should methodically discuss cause-and-effect relationships between events.
 - Mnemonic devices should be taught to the client to enhance medium-term delayed memory.
 - He should learn to identify moments when he begins to feel overwhelmed or confused, so that he can respond appropriately.
4. His *cognitive-behavioral therapist* can help the client in the following ways:
- As with his learning specialist, he should learn to identify moments when he begins to feel overwhelmed or confused, so that he can respond appropriately.
 - He should figure out ways to lessen the effect of him shutting down. For example, if he can identify moments in which he has shut down, he can learn to take breaks, detour to a different activity (even if for a short amount of time), or learn cognitive strategies to combat it.
 - He should be encouraged to find appropriate ways to convey to teachers that he does not understand a concept, minimizing embarrassment or public attention.
 - He should be encouraged, along with his parents, to create a schedule and plan for doing his homework each night that allows for both structure and flexibility for breaks. This plan should include ways of decreasing sensory stimulation during the period he is doing his homework.
5. While it is likely that he will perform adequately in most school environments that can be sensitive to his specific needs, he will likely thrive in an environment that can provide him with a more *individualized, tailored educational program*. One option is XXXXXX, which provides highly individualized, personal educational programming for middle and high school students. Another option that would be a good match for his learning needs is XXXXXX, which also offers highly individualized educational programming.

A. Jordan Wright, PhD, ABAP
New York State License: XXXXXX

Mr. S.
100 Main Street
Smalltown, TX 7XXXX

Dear Mr. S.

This is the letter I promised you summarizing the results of the psychological assessment we did together this past month. As I did in our last session, I'll structure the letter by addressing each of your questions for the assessment:

Why can't I cry?

Mr. S., as you may remember, we worked mainly with two major personality tests, the Minnesota Multiphasic Personality Inventory—2 (MMPI-2)—the long true-false test you took—and the Rorschach inkblots. The combination of your scores on these two tests helped me understand why you can't cry at those times you wish you could.

The MMPI-2 excels at measuring people's outside "persona"—the way they present themselves to the world and generally think about themselves. Your results showed that you have *very* strong psychological coping mechanisms and that these help you keep difficult feelings "under wraps" and out of awareness. I believe these skills led to your reporting much less depression, anger, and anxiety than do most people who fill out the MMPI-2. This "picture" of you from the MMPI-2 is likely very similar to how you and others usually think of you—as a sturdy, nonreactive guy who is slow to anger, resilient, and not easily upset. This view of you captures important truths about who you are.

In contrast, the Rorschach excels at revealing "underlying" feelings—those that are affecting us at some level but of which we may be unaware. On this test you scored very much like people who are dealing with severely painful feelings, including sadness, anxiety, anger, remorse, alienation, and self-doubt. I suspect these feelings also capture an important reality about you—that is, that you carry a lot of pain inside that is left over from difficult events earlier in your life. By necessity, you found a way to put this pain aside when it first arose. Now it feels threatening to get too in touch with any feelings related to these "stored" emotions, so they're in "deep storage" and you're unable to cry even when you feel upset.

The Rorschach also suggests another factor in your not crying: your emotional "controls" appear to be less developed than would be optimal for your age. As you so aptly said, at this point you have more of an "on-off" switch for emotions rather than a "rheostat." This situation is common when one has had to shut off and avoid feelings early in life in order to be safe. While this

coping strategy served you well, your emotional management skills didn't get to grow up as fast as the rest of you. Now, if you open up inside to strong emotions, they are likely to flood you, overwhelm you, disrupt your ability to think clearly, and leave you feeling confused and out of control. Thus, you generally keep the switch turned "off."

So, in short, until you've had a chance to develop better emotional controls and to find support from someone who is knowledgeable about emotional blockages, it may be too scary to cry. In the meantime, it does seem that you are developing more access to your feelings and finding that this helps you rather than hurts you (e.g., in intimate relationships).

Why do I remember so little of my childhood?

Mr. S., as you know, for your assessment I talked with Dr. Smith (your therapist from ages 7–14), and he told me his impressions of your childhood. I also know some things about your family situation from talking to your parents when I assessed them and your sister. Last, you did write about some early events on the Early Memory Procedure. (I'm returning the original booklet to you with this letter.)

From all accounts, when you were little, you experienced some traumatic abandonments and violent family scenes that would have been quite emotionally overwhelming for any small child. On top of this, you didn't have adequate emotional support at that time from those around you so you could process these experiences and the feelings they must have generated. So at least some of the time, you reportedly "tuned out" or "dissociated" as a coping mechanism. This is very much like an emotional "fuse" blowing to shut down certain parts of the brain when it is flooded and overwhelmed. Typically this is a last-ditch survival mechanism when one is unable to physically flee or fight in a terrifying situation.

What all this means is that some of your childhood memories were probably not even "encoded" at the time, because you were so overwhelmed and in a state of emotional shock. (Shock is an altered state where our senses screen out incoming data to keep us from getting more overwhelmed.) In addition, it's likely that some memories are still there, but they are currently being kept out of awareness because you aren't yet ready to face the feelings that would result if you had those memories back. Such feelings could overwhelm you even now until you develop better emotional controls. Plus, as you yourself said, you're not ready to deal with the effects on your relationship with your parents of knowing all that went on when you were a child. Thus, you have been intuitively taking care of yourself by not remembering much from your childhood.

Why did I repeatedly put myself in the role of a caregiver with girlfriends and women friends?

Let me remind you of some things you already know: You and your father have been caregivers in the past of your mother, so this role was familiar to you and modeled to you as the way to be with women. Also, the caregiver role has been a way to feel good about yourself; there is a part of you that genuinely

wants to help people and feels good when you can, and such behavior has been rewarded in the past. Third, from watching your mother, you decided you never wanted to lean too much on others; this pushed you to “overcorrect”—that is, to deny your own needs to be cared for and to focus mainly on caring for others.

The testing adds the following pieces to the puzzle: Because your feelings were shut down to protect you, you weren’t very in touch with your anger and thus couldn’t realize that at times you felt taken advantage of by women who were leaning on you heavily. Also, for many of us, when we are focused outside ourselves on helping other people with their emotional pain, it helps keep our own pain at bay. Last, I believe all this is a way we go about healing also. We project our own sad and hurt feelings out on others (or find people who seem to embody these feelings) and then we try to “fix” the problem outside ourselves. Eventually, if all goes well, we eventually are forced into a major confrontation with ourselves—that we too have needs for support and caring—and that we can’t address those needs by simply caring for others. You seem to have been having such a realization lately.

It’s so great to see the progress you’ve made in this area, Mr. S. The testing says you are still at risk to fall into caregiver roles, so you’ll want to keep an eye out for this and pay attention to any anger you feel about supporting others. Such anger could be a sign that you are doing too much.

Why do I often not know how I am feeling?

Mr. S., you described a shift around ninth grade when you started to feel more in control and more sure of yourself. I wonder if this is the time when you finally achieved some success in shutting down your own feelings so you would not feel anxious, stirred up, and overwhelmed all the time. However, you were probably already using pieces of this coping mechanism much earlier. Again, I believe such a strategy would have helped you not “rock the boat” at home—which would have helped keep your mom stable and your parents from fighting. Your coping style was modeled after someone you admired and were close to—your dad. And as we talked about, as a child you also got rewards from other people when you didn’t show normal feelings of anger, impatience, or irritation—for being so “adult,” “precocious,” and “well mannered.”

Last, what I understand from the testing is this: After a while, the feelings you had put aside were so huge and painful that you would need help accessing them and managing them. Plus, experience had taught you that the people you loved most—your parents—were not really able to help you process painful emotions. This left you with few choices. Since you, like the rest of us, don’t like feeling out of control of your emotions, you did what made sense: kept your feelings out of awareness. As we discussed, I believe that when you are ready, it will be very helpful to “open up” some of these feelings—with the support of a therapist—and that doing so will help you succeed in your intimate relationships.

Mr S., I very much enjoyed working with you and I hope the assessment and this letter are helpful. Please don’t hesitate to call, visit, or e-mail me if you have any questions about the results of your assessment.

Last, I have a request for you. Would you be willing to fill out the enclosed forms giving me feedback about the assessment and mail them back to me? Your honest comments would help me serve other people in the future.

With my very best wishes for your future,
Stephen E. Finn, Ph.D.
Licensed Psychologist

Source note: Report written and submitted by Stephen E. Finn, PhD, Center for Therapeutic Assessment.

RECOMMENDED READING

- Allyn, J. B. (2012). *Writing to clients and referring professionals about psychological assessment results: A handbook of style and grammar*. New York, NY: Routledge.
- Finn, S. E. (2007). *In our client's shoes: Theory and techniques of therapeutic assessment*. Mahwah, NJ: Erlbaum.
- Groth-Marnat, G., & Davis, A. (2014). *Psychological report writing assistant*. Hoboken, NJ: Wiley.
- Harvey, V. S. (2006). Variables affecting the clarity of reports. *Journal of Clinical Psychology*, 62, 5–18.
- Wright, A. J. (2010). *Conducting psychological assessment: A guide for practitioners*. Hoboken, NJ: Wiley.

TEST PUBLISHERS/DISTRIBUTORS

American Guidance Service, Inc. (see Pearson Assessment)

Consulting Psychologists Press

1055 Joaquin Road, 2nd Floor

Mountain View, CA 94043

1-650-969-8901

1-800- 624-1765

Fax: 1-650-969-8608

www.cpp.com

List includes: California Psychological Inventory, Fundamental Interpersonal Relations Orientation—B, Myers Briggs Type Indicator, Strong Interest Inventory

Harcourt Assessment (see Pearson Assessment)

Institute for Personality and Ability Testing

1801 Woodfield Drive

Savoy, IL 61874

1-800-225-4728

1-217-352-4739

Fax: 1-217-352-9674

www.ipat.com/

List includes: Sixteen Personality Factors (16PF)

Lafayette Instrument Company

P.O. Box 5729

Lafayette, IN 47903

1-800-428-7545

www.lafayetteinstrument.com

List includes: Hand dynamometer (grip strength)

NFER-Nelson Publishing Co.

Darville House

2 Oxford Road

East Windsor

Berkshire 21A IDF, UK

www.nfer.ac.uk

List includes: National Adult Reading Test

National Computer Systems, Inc. (NCS; see Pearson Assessment)

Pearson Assessment (previously American Guidance Service, Inc., Harcourt Assessment, Psychological Corporation, and National Computer Systems)

P.O. Box 599700

San Antonio, TX 78259

1-800-211-8378

Fax: 1-800-232-1223

List includes: Beck Depression Inventory—II, Bender Visual Motor Gestalt—II, Brief Symptom Inventory, California Verbal Learning Test, Career Assessment Inventory, Children's Depression Inventory, Children's Category Test, Children's Memory Scale, Kaufman Assessment Battery for Children—II, Kaufman Adolescent and Adult Intelligence Test, Kaufman Brief Intelligence Test, Kaufman Test of Educational Achievement—III, Millon Behavioral Health Inventory, Millon Clinical Multiaxial Inventory—IV, Minnesota Multiphasic Personality Inventory—2, Minnesota Multiphasic Personality Inventory—2-RF, Paced Auditory Serial Addition Test, Repeatable Battery for the Assessment of Neuropsychological Status Update, Stroop Color-Word Test, Symptom Checklist 90-R, Taylor Johnson Temperament Analysis, Test of Memory Malingering, Wechsler Adult Intelligence Scale—IV, Wechsler Individual Achievement Test—II, Wechsler Intelligence Scale for Children—V, Wechsler Preschool and Primary Scale for Children—IV, Wechsler Memory Scale—IV, Wechsler Test of Adult Reading

Psychological Assessment Resources

16204 N. Florida Avenue

Lutz, FL 33549

1-800-331-8378

Fax: 1-800-727-9329

Tech Support: 1-800-899-8378

www.parinc.com

List includes: BarOn Emotional Quotient Inventory, Behavioral Assessment of the Dysexecutive Syndrome, Bender Visual Motor Gestalt Test—2, Benton Visual Retention Test, Boston Diagnostic Aphasia Examination, Boston Naming Test, Category Test (computer version), Children's Apperception Test, Children's Auditory Verbal Learning Test, Children's Category Test, Cognitive Assessment System, Color Trails Test, Connor's Rating Scales, Finger Tapping, Hand Dynamometer, Hare Psychopathy Checklist—Revised, House Tree Person, KOPPITZ-2 (Koppitz Developmental Scoring System for the Bender), NEO-PI-3, Personality Assessment Inventory, Personality Disorder Interview—IV, Rey Auditory Verbal Learning Test, Rey Complex Figure and Recognition Trial, Rivermead Behavioural Memory Test, Rorschach Comprehensive System, Self Directed Search, Sentence Completion Series, State Trait Anger Expression Inventory, State Trait Anxiety Inventory, State Trait Anxiety Inventory for Children, Stroop Neuropsychological Screening Test, Stroop Color and Word Test, Tactual Performance Test, Taylor Johnson Temperament Analysis, Test of Everyday Attention, Test of Nonverbal Intelligence, Thematic

Apperception Test, Wide Range Achievement Test—4, Wide Range Assessment of Memory and Learning, Wisconsin Card Sorting Test

Psychological Corporation (see Pearson Assessment)

Reitan Neuropsychology Laboratory

2920 South 4th Avenue

Tucson, AZ 85713-4819

rn-lab.net

List includes: Halstead Reitan Neuropsychological Test Battery, Neuropsychological History Questionnaire

Riverside Publishing Co.

8420 Bryn Mawr Avenue

Chicago, IL 60631

1-201-729-6031

www.riverpub.com

List includes: Bender Visual Motor Gestalt Test-II, Das Naglieri Cognitive Assessment System, Stanford-Binet-5, Woodcock Johnson-IV

Rorschach Performance Assessment System, Inc.

www.r-pas.org

List includes: Rorschach Performance Assessment System (R-PAS)

Western Psychological Services

12031 Wilshire Boulevard

Los Angeles, CA 90025-1251

1-800-222-2670

www.wpspublish.com

List includes: AAMR Adaptive Behavior Scales, Adaptive Behavior Assessment Scale—III, Adolescent Apperception Test, Behavior Rating Inventory of Executive Functions, Bender Visual Motor Gestalt Test—2, Children's Category Test, Children's Depression Inventory, Comprehensive Test of Nonverbal Intelligence, Connor's Rating Scales, Draw-a-Person, Eating Disorders Inventory, House-Tree-Person, Human Figure Drawing Test, Family Apperception Test, Kaufman Adolescent and Adult Intelligence Test, Kaufman Assessment Battery for Children, Luria Nebraska Neuropsychological Battery, Millon Index of Personality Styles, Personality Assessment Inventory, Personality Inventory for Children, Psychopathy Checklist, Rey Auditory and Verbal Learning Test, Rivermead Behavioural Memory Test, Rey Auditory Verbal Learning Test, Roberts Apperception Test for Children, Self-Directed Search, State Trait Anger Expression Inventory, Symbol Digit Modalities Test, Taylor Johnson Temperament Analysis, Test of Everyday Attention, Tell Me A Story, Wide Range Achievement Test—4, Wide Range Test of Memory and Learning, Wisconsin Card Sorting

TESTING ORGANIZATIONS

American Board of Assessment Psychology

1581 Brickell Avenue, #106
Miami, FL 33129
1-305-372-0010 ext. 207
Fax: 1-305-374-8378
www.assessmentpsychologyboard.org

American Psychological Association

750 First Street, NE
Washington, DC 20002-4242
1-800-374-2721
1-202-336-5500
www.apa.org

Board on Testing and Assessment

The National Academies
500 Fifth Street, NW, 11th Floor
Washington, DC 20001
1-202-334-2000
Fax: 1-202-344-1294
www.nationalacademies.org/bota/

Buros Institute of Mental Measurements

University of Nebraska, Lincoln
21 Teachers College Hall
Lincoln, NE 68588-0348
1-402-472-6203
Fax: 1-402-472-6207
www.buros.org

International Neuropsychological Society

700 Ackerman Road, Suite 625
Columbus, Ohio 43202
1-614-263-4200
Fax: 1-614- 263-4366
www.the-ins.org/

National Academy of Neuropsychology

7555 East Hampden Avenue, Suite 525

Denver, CO 80231

1-303-691-3694

Fax: 1-303-691-5983

www.nanonline.org

National Council on Measurement in Education (NCME)

Central Office

100 North 20th Street, Suite 400

Philadelphia, PA 19103

1-215-461-6263

Fax: 1-215-564-2175

www.ncme.org

Society for Industrial and Organizational Psychology

440 East Poe Road, Suite 101

Bowling Green, OH 43402

1-419-353-0032

Fax: 1-419-352-2645

www.siop.org/contact.aspx

Society for Personality Assessment

6109H Arlington Boulevard

Falls Church, Virginia 22044

1-703-534-4SPA (1-703-534-4772)

Fax: 1-703-534-6905

www.personality.org

References

- Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the child behavior checklist and revised behavior profile*. Burlington: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Acker, M. B., & Davis, J. R. (1989). Psychology test scores associated with late outcome head injury. *Neuropsychology*, 3, 1–10.
- Ackerman, M. J. (2006a). *Clinician's guide to child custody evaluations* (3rd ed.). Hoboken, NJ: Wiley.
- Ackerman, M. J. (2006b). Forensic report writing. *Journal of Clinical Psychology*, 62, 59–72.
- Ackerman, S. J., Hilsenroth, M. J., Baity, M. R., & Blagys, M. D. (2000). Interaction of therapeutic process and alliance during psychological assessment. *Journal of Personality Assessment*, 75, 82–109.
- Acklin, M. W., & Bernat, E. (1987). Depression, alexithymia, and pain prone disorder: A Rorschach study. *Journal of Personality Assessment*, 51, 462–479.
- Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment*, 74, 15–47.
- Adams, K. M. (1980). In search of Luria's battery: A false start. *Journal of Consulting and Clinical Psychology*, 48, 511–516.
- Aegisdottir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist*, 34, 341–382.
- Agbayewa, M. O. (1990). An exploratory use of the Symptoms Checklist-90 in a mixed geriatric study group. *Journal of the American Geriatrics Society*, 38(7), 773–776.
- Ahava, G. W., Iannone, C., Grebstein, L., & Schirling, J. (1998). Is the Beck Depression Inventory reliable over time? An evaluation of multiple test-retest reliability in a nonclinical college student sample. *Journal of Personality Assessment*, 70, 222–231.
- Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology*, 48, 251–257.
- Aiken, L. R., & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston, MA: Pearson Education.
- Alamilla, S. G., & Wojcik, J. V. (2013). Assessing for personality disorders in the Hispanic client. In L. T. Benuto (Ed.), *Guide to psychological assessment with Hispanics* (pp. 215–241). New York, NY: Springer.
- Albert, S., Fox, H. M., & Kahn, M. W. (1980). Faking psychosis on the Rorschach: Can expert judges detect malingering? *Journal of Personality Assessment*, 44, 115–119.
- Alfonso, V. C., Johnson, A., Patinella, L., & Radar, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. *Psychology in the Schools*, 35, 119–125.
- Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment*, 7(2), 119–129.

- Allen, J. J. (1998). Personality assessment with American Indians and Alaska Natives: Instrument considerations and service delivery style. *Journal of Personality Assessment*, 70, 17–42.
- Allen, J. J., & Dana, R. H. (2004). Methodological issues in cross-cultural and multicultural Rorschach research. *Journal of Personality Assessment*, 82(2), 189–208.
- Alloy, L. B., Abramson, W. G., Whitehouse, W. G., Hogan, M. E., Tashman, N. A., Steinberg, D. L., ... & Donovan, P. (1999). Depressogenic cognitive styles: Predictive validity, information processing, and personality characteristics, and developmental origins. *Behavior Research and Therapy*, 37, 503–531.
- Allyn, J. B. (2012). *Writing to clients and referring professionals about psychological assessment results: A handbook of style and grammar*. New York, NY: Routledge.
- Almagor, M., Koren, D. (2001). The adequacy of the MMPI-2 Harris-Lingoes subscales: A cross-cultural factor analytic study of scales D, Hy, Pd, Sc, and Ma. *Psychological Assessment*, 13(2), 199–215.
- Al-Musawi, N. M. M. (2001). Psychometric properties of the Beck Depression Inventory–II with university students in Bahrain. *Journal of Personality Assessment*, 77(3), 568–579.
- Alnacs, R., & Torgerson, S. (1989). Clinical differentiation between major depression only, major depression with panic disorder, and panic disorder: Childhood personality, and personality disorder. *Acta Psychiatrica Scandinavica*, 79, 370–377.
- Alterman, A. I., Zaballero, A. R., Lin, M. M., Siddiqui, N., Brown, L. S., Rutherford, M. J., & McDermott, P. A. (1995). Personality Assessment Inventory (PAI) scores of lower-socioeconomic African American and Latino methadone maintenance patients. *Assessment*, 2(1), 91–100.
- Amabile, T. M. (1983). *The social psychology of creativity*. New York, NY: Springer-Verlag.
- Ambrosini, P. J. (2000). Historical development and present status of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS). *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 49–58.
- Ambrosini, P. J., Metz, C., Prabucki, K., & Lee, J. (1989). Videotape reliability of the third revised edition of K-SADS. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 723–728.
- American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Journal of Managed Care. (1999). Introduction. *American Journal of Managed Care*, 5, S764–S766.
- American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- American Psychological Association. (1987). General guidelines for providers of psychological services. *American Psychologist*, 42, 7.
- American Psychological Association. (1991). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1075.
- American Psychological Association. (2010). Guidelines for child custody evaluations in family law proceedings. *American Psychologist*, 65(9), 863–867.
- American Psychological Association Division 12—Society of Clinical Psychology. (2009). Research-supported psychological treatments. Retrieved from www.PsychologicalTreatments.org
- Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. (1973). *Rorschach responses in old age* (rev. ed.). New York, NY: Brunner/Mazel.
- Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. (1974). *Child Rorschach responses: Developmental trends from two to ten years* (rev. ed.). New York, NY: Brunner/Mazel.
- Anastopoulos, A. D., Spisto, M. A., & Maher, M. C. (1994). The WISC-III Freedom from Distractibility Factor: Its utility in identifying children with attention deficit hyperactivity disorder. *Psychological Assessment*, 6, 368–371.
- Anderson, E. M., & Lambert, M. J. (1995). Short-term dynamically oriented psychotherapy: A review and meta-analysis. *Clinical Psychology Review*, 15, 503–514.
- Anderson, K. E., & Savage, C. R. (2004). Cognitive and neurobiological findings in obsessive-compulsive disorder. *Psychiatric Clinics of North America*, 27, 37–47.
- Andrews, H. B. (2001). Back to basics: Psychotherapy is an interpersonal process. *Australian Psychologist*, 36, 107–114.
- Anestis, J. C., Finn, J. A., Gottfried, E., Arbisi, P. A., & Joiner, T. E. (2015). Reading the road signs: The utility of the MMPI-2 Restructured Form validity scales in the prediction of premature termination. *Assessment*, 22(3), 279–288.
- Anestis, J. C., Gottfried, E. D., & Joiner, T. E. (2015). The utility of MMPI-2-RF substantive scales in prediction of negative treatment outcomes in a community mental health center. *Assessment*, 22(1), 23–35.
- Anker, M.G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, 77(4), 693–704.
- Antonuccio, D. O., Danton, W. G., & DeNelsky, G. Y. (1995). Psychotherapy versus medication for depression: Challenging the conventional wisdom with data. *Professional Psychology: Research and Practice*, 26, 574–585.
- Antony, M. M., & Barlow, D. (Eds.). (2011). *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders* (2nd ed.). New York, NY: Guilford Press.
- Apter, A., Bleich, A., Plutchik, R., Mendelsohn, S., & Tyrano, S. (1988). Suicidal behavior, depression, and conduct disorder in hospitalized adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 696–699.
- Aram, D. M., & Ekelman, B. L. (1986). Cognitive profiles of children with early onset unilateral lesions. *Developmental Neuropsychology*, 2, 155–172.
- Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency Psychopathology Scale, F(p). *Psychological Assessment*, 7, 424–431.
- Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2003). Empirical correlates of common MMPI-2 two-point codes in male psychiatric inpatients. *Assessment*, 10(3), 237–247.
- Archer, R. P. (1984). Use of the MMPI with adolescents: A review of salient issues. *Clinical Psychology Review*, 4, 241–251.
- Archer, R. P. (1987). *Using the MMPI with adolescents*. Hillsdale, NJ: Erlbaum.
- Archer, R. P. (1992a). *MMPI-A: Assessing adolescent psychopathology*. Hillsdale, NJ: Erlbaum.
- Archer, R. P. (1992b). Review of the Minnesota Multiphasic Personality Inventory–2. In J. J. Kramer & J. C. Conoley (Eds.), *Eleventh mental measurements yearbook* (pp. 558–562). Lincoln, NE: Buros Institute of Mental Measurements.
- Archer, R. P. (2005). *MMPI-A: Assessing adolescent psychopathology* (3rd ed.). Hillsdale, NJ: Erlbaum.

- Archer, R. P. (2006). *Forensic uses of clinical assessment instruments*. Mahwah, NJ: Erlbaum, 2006.
- Archer, R. P., Buffington-Vollum, Stredny, J. K., & Handel, R. W. (2006). A survey of test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 64–94.
- Archer, R. P., Griffin, R., & Aiduk, R. (1995). MMPI-2 clinical correlates for ten common codes. *Journal of Personality Assessment*, 65, 391–407.
- Archer, R. P., & Jacobson, J. M. (1993). Are critical items “critical” for the MMPI-A? *Journal of Personality Assessment*, 61, 547–556.
- Archer, R. P., & Krishnamurthy, R. (1997a). MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment*, 69, 517–533.
- Archer, R. P., Maruish, M. E., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice*, 22, 247–252.
- Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment*, 7, 227–236.
- Arizmendi, T. G., Beutler, L. E., Shanfield, S., Crago, M., & Hagaman, R. (1985). Client-therapist similarity and psychotherapy outcome: A microscopic approach. *Psychotherapy: Theory, Research, and Practice*, 22, 16–21.
- Armbruster, G. L., Miller, A. S., & Exner, J. E. (1974). Rorschach responses of parachute trainees at the beginning of training and shortly before their first jump. Rorschach Workshops (Study No. 201, unpublished).
- Armengol, C. G. (2001). The referral process. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 47–60). Lutz, FL: Psychological Assessment Resources.
- Armengol, C. G., Moes, E. J., Penney, D. L., & Sapienza, M. M. (2001). *Writing client-centered recommendations*. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 141–160). Lutz, FL: Psychological Assessment Resources.
- Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington, DC: War Department, Adjutant General's Office.
- Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory–II with primary care medical patients. *Health Psychology*, 20, 112–119.
- Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281–322.
- Asaad, G. (2000). Somatization disorder. In M. Hersen & M. Biaggio (Eds.), *Effective brief therapies: A clinician's guide* (pp. 179–190). San Diego, CA: Academic Press.
- Association of Test Publishers. (2000). *Guidelines for computer-based testing*. Washington, D.C.: Author.
- Atkinson, L., Quarington, B., Alp, I. E., & Cyr, J. J. (1986). Rorschach validity: An empirical approach to the literature. *Journal of Clinical Psychology*, 42, 360–362.
- Aue, T., Lavelle, L. A., & Cacioppo, J. T. (2009). Great expectations: What can fMRI research tell us about psychological phenomena? *International Journal of Psychophysiology*, 73(1), 10–16.
- Austin, A., & Craig, S. L. (2015). Empirically supported interventions for sexual and gender minority youth. *Journal of Evidence-Informed Social Work*, (ahead-of-print), 1–12.
- Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, 15, 433–442.
- Axelrod, B. N. (2001). Administration duration for the Wechsler Adult Intelligence Scale-III and Wechsler Memory Scale-III. *Archives of Clinical Neuropsychology*, 16, 293–301.
- Axelrod, B. N., Ryan, J. J., & Woodward, J. L. (2001). Cross-validation of prediction equations for Wechsler Memory Scale-III. *Assessment*, 8, 367–372.

- Axelrod, B. N., Vanderploeg, R. D., & Schinka, J. A. (1999). Comparing methods for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology*, 14, 341–346.
- Axelrod, B. N., & Woodward, J. L. (2000). Parsimonious prediction of Wechsler Memory Scale–III memory indices. *Psychological Assessment*, 12, 431–435.
- Baade, L. E., & Schoenberg, M. A. (2004). A proposed method to estimate premorbid intelligence utilizing group achievement measures from school records. *Archives of Clinical Neuropsychology*, 19, 227–244.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839.
- Baddeley, A. D., Kopelman, M. D., & Wilson, B. A. (2002). *The handbook of memory disorders* (2nd ed.). New York, NY: Wiley.
- Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment*, 72, 308–320.
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment*, 14(1), 16–26.
- Baer, R. A., Wetter, M. W., & Berry, D. T. (1995). Effects of information about validity scales underreporting symptoms on the MMPI-2: An analogue investigation. *Assessment*, 2, 189–200.
- Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review*, 12, 509–525.
- Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment*, 7, 419–423.
- Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, 7, 84–92.
- Bagby, R. M., Gillis, J. R., & Rogers, R. (1991). Effectiveness of the Millon Clinical Multiaxial Inventory Validity Index in the detection of random responding. *Psychological Assessment*, 3, 285–287.
- Bagby, R. M., Gillis, J. R., Toner, B. B., & Goldberg, J. (1991). Detecting fake-good and fake-bad responding on the Millon Clinical Multiaxial Inventory-II. *Psychological Assessment*, 3, 496–498.
- Bagby, R. M., & Marshall, M. B. (2004). Assessing underreporting response bias on the MMPI-2. *Assessment*, 11(2), 115–126.
- Bagby, R. M., Marshall, M. B., Basso, M. R., Nicholson, R. A., Bacchiochi, J., & Miller, L. S. (2005). Distinguishing bipolar depression, major depression, and schizophrenia with the MMPI-2 Clinical and Content Scales. *Journal of Personality Assessment*, 84, 89–95.
- Bagby, R. M., Rogers, R., Buis, T., & Kalembe, V. (1994). Malingered and defensive response styles on the MMPI-2: An examination of validity scales. *Assessment*, 1(1), 31–38.
- Bagby, R. M., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of defensive responding in clinical and nonclinical samples. *Psychological Assessment*, 9, 406–413.
- Bakker, C., Bakker-Rabaud, M., & Breit, S. (1978). The measurement of assertiveness and aggressiveness. *Journal of Personality Assessment*, 42, 277–284.
- Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach Depression Indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment*, 57, 465–476.
- Ball, J. D., Archer, R. P., & Imhof, E. A. (1994). Time requirements of psychological testing: A survey of practitioners. *Journal of Personality Assessment*, 63, 239–249.
- Ballenger, J. F., Caldwell-Andrews, A., & Baer, R. A. (2001). Effects of positive impression management on the NEO Personality Inventory-Revised in a clinical population. *Psychological Assessment*, 13(2), 254–260.

- Bamgbose, O., Smith, G. T., Jesse, R. C., & Groth-Marnat, G. (1980). A survey of the current and future directions of professional psychology in acute general hospitals. *Clinical Psychologist*, 33, 24–25.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC scaled scores. *Journal of Learning Disabilities*, 7, 272–273.
- Barber, T. X., & Silver, M. J. (1968). Fact, fiction and the experimenter bias effect. *Psychological Bulletin Monograph Supplement*, 70, 1–29.
- Barlow, D. H. (1988). *Anxiety and its disorders: The nature and treatment of anxiety and panic*. New York, NY: Guilford Press.
- Barlow, D. H. (Ed.). (2014). *Clinical handbook of psychological disorders: A step-by-step treatment manual* (5th ed.). New York, NY: Guilford Press.
- Barlow, D. H., Conklin, L. R., & Bentley, K. H. (2015). Psychological treatments for panic disorders, phobias, and social and generalized anxiety disorder. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (4th ed., pp. 409–462). New York, NY: Oxford University Press.
- Barlow, D. H., Craske, M. G., Cerny, J. A., & Klosko, J. S. (1989). Behavioral treatment of panic disorder. *Behavior Therapy*, 20, 261–282.
- Bar-On, R. (1998). *Bar-On Emotional Quotient Inventory (EQ-itm): Technical manual*. Toronto, Canada: Multi Health Systems.
- Barona, A., Reynolds, C., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*, 26, 74–75.
- Barrera, M., & González Castro, F. G. (2006). A heuristic framework for the cultural adaptation of interventions. *Clinical Psychology: Science and Practice*, 13(4), 311–316.
- Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Incremental validity of the MMPI-2 content scales in an outpatient mental health setting. *Psychological Assessment*, 11, 39–47.
- Bartholomew, D. J. (2006). *Measuring intelligence: Facts and fallacies*. Cambridge, UK: Cambridge University Press.
- Bartlett, J. A., Schleifer, S. J., Johnson, R. L., & Keller, S. E. (1991). Depression in inner city adolescents attending an adolescent medicine clinic. *Journal of Adolescent Health*, 12, 316–318.
- Basham, R. B. (1992). Clinical utility of the MMPI research scales in the assessment of adolescent acting out. *Psychological Assessment*, 4, 483–492.
- Basso, M. B., Harrington, K., Matson, M., & Lowery, N. (2000). Sex differences on the WMS-III: Findings concerning verbal paired associates and faces. *Clinical Neuropsychologist*, 14, 231–235.
- Bateman, A. W., & Fonagy, P. (Eds.) (2012). *Handbook of mentalizing in mental health practice*. Arlington, VA: American Psychiatric Association.
- Baucom, D. H., Epstein, N., Rankin, L. A., & Burnett, C. K. (1996). Assessing relationship standards: The Inventory of Specific Relationship Standards. *Journal of Family Psychology*, 10, 72–88.
- Baucom, D. H., Sayers, S. L., & Duhe, A. (1989). Marital attributions: Issues concerning attributional pattern and attributional style. *Journal of Personality and Social Psychology*, 56, 596–607.
- Bauer, R. M. (2000). The flexible battery approach to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 419–448). Hillsdale, NJ: Erlbaum.
- Beatty, W. W., Mold, J. W., & Gontkovsky, S. T. (2003). RBANS performance: Influence of sex and education. *Journal of Clinical and Experimental Neuropsychology*, 25, 1065–1069.
- Beatty, W. W., Ryder, K. A., Gontkovsky, S. T., Scott, J. G., McSwan, K. L., & Bharucha, K. J. (2003). Analyzing the subcortical dementia syndrome of Parkinson's disease using the RBANS. *Archives of Clinical Neuropsychology*, 18, 509–520.

- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York, NY: Guilford Press.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II manual*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Steer, R. A., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.
- Beck, J. G. (1994). *Cognitive therapy: Basics and beyond*. New York, NY: Guilford Press.
- Beck, J. G., & Zebb, B. J. (1994). Behavioral assessment and treatment of panic disorder: Current status, future directions. *Behavior Therapy*, 25, 581–611.
- Beck, S. J. (1937). *Introduction to the Rorschach method: A manual of personality study*. American Orthopsychiatric Association Monograph, 1.
- Beck, S. J. (1945). *Rorschach's test: A variety of personality pictures* (Vol. 2). New York, NY: Grune & Stratton.
- Beck, S. J. (1960). *The Rorschach experiment*. New York, NY: Grune & Stratton.
- Beck, S. J. (1961). *Rorschach's test: Basic processes* (Vol. 1). New York, NY: Grune & Stratton.
- Beck, S. J. (1968). Reality, Rorschach and perceptual theory. In A. I. Rabin (Ed.), *Projective techniques in personality assessment* (pp. 115–135). New York, NY: Springer.
- Beebe, D. W., Pfiffner, L. J., & McBurnett, K. (2000). Evaluation of the validity of the Wechsler Intelligence Scale for Children–3rd ed. Comprehension and Picture Arrangement subtests as measures of social intelligence. *Psychological Assessment*, 12, 97–101.
- Behnke, S. (2004, November). Ethics rounds: Release of test data and the new ethics code. *Monitor*, 35, 90–91.
- Beier, E. G. (1966). *The silent language of psychotherapy*. New York, NY: Aldine.
- Bellack, A. S., & Hersen, M. (Eds.). (1976). *Behavioral assessment: A practical handbook*. New York, NY: Pergamon Press.
- Bellack, A. S., & Hersen, M. (Eds.). (1988). *Behavioral assessment: A practical handbook* (3rd ed.). New York, NY: Pergamon Press.
- Bellack, A. S., & Hersen, M. (Eds.). (1998). *Behavioral assessment: A practical handbook* (4th ed.). New York, NY: Pergamon Press.
- Bellack, A. S., Hersen, M., & Turner, S. M. (1979). Relationship of role playing and knowledge of appropriate behavior to assertion in the natural environment. *Journal of Consulting and Clinical Psychology*, 47, 670–678.
- Beltrán, M. D. C., Freyre, M. Á., & Hernández-Guzmán, L. (2012). El Inventario de Depresión de Beck: Su validez en población adolescente. *Terapia psicológica*, 30(1), 5–13.
- Bender, L. (1938). *A visual motor gestalt test and its clinical uses* (Research Monographs No. 3). New York: American Orthopsychiatric Association.
- Benjamin, A. B., Mossman, D., Graves, N. S., & Sanders, R. D. (2006). Tests of a symptom checklist to screen for comorbid psychiatric disorders in alcoholism. *Comprehensive Psychiatry*, 47, 227–233.
- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y. S., & Butcher, J. N. (1989). The psychometric stability of rewritten MMPI items. *Journal of Personality Assessment*, 53, 645–653.
- Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI-2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment*, 3, 634–640.
- Ben-Porath, Y. S., Hostetler, K., Butcher, J. N., & Graham, J. R. (1989). New subscales for the MMPI-2 Social Introversion (Si) Scale. *Psychological Assessment*, 1, 169–174.

- Ben-Porath, Y. S., McCully, E., & Almagor, M. (1993). Incremental validity of the MMPI-2 content scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment*, 61, 557–575.
- Ben-Porath, Y. S., Shondrick, D. D., & Stafford, K. P. (1995). MMPI-2 and race in a forensic diagnostic sample. *Criminal Justice and Behavior*, 22, 19–32.
- Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *The Minnesota Multiphasic Personality Inventory–2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y. S., & Waller, N. G. (1992). “Normal” personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment*, 4(1), 14–19.
- Benson, N. F. (2014). Review of NEO Personality Inventory–3. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook* (pp. 478–481). Lincoln, NE: Buros Center for Testing.
- Benton, A. L. (1974). *The Revised Visual Retention Test* (4th ed.). New York, NY: Psychological Corporation.
- Benton, A. L. (1979). Visuoceptive, visuospatial, and visuoconstructive disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 186–232). New York, NY: Oxford University Press.
- Benton, A. L., & Hamsher, K. (1989). *Multilingual Aphasia Examination* (2nd ed.). Iowa City, IA: AJA Associates.
- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 217–270). New York, NY: Wiley.
- Bergua, V., Meillon, C., Potvin, O., Bouisson, J., Le Goff, M., Rouaud, O., ... Amieva, H. (2012). The STAI-Y trait scale: Psychometric properties and normative data from a large population-based study of elderly people. *International Psychogeriatrics*, 24(7), 1163–1171.
- Bernert, R. A., Hom, M. A., & Roberts, L. W. (2014). A review of multidisciplinary clinical practice guidelines in suicide prevention: Toward an emerging standard in suicide risk assessment and management, training and practice. *Academic Psychiatry*, 38(5), 585–592.
- Berry, D. T., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analysis. *Journal of Personality Assessment*, 11, 585–598.
- Berzins, J. I. (1977). Therapist-patient matching. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research* (pp. 222–251). New York, NY: Pergamon Press.
- Beutler, L. E. (1983). *Eclectic psychotherapy: A systematic approach*. New York, NY: Pergamon Press.
- Beutler, L. E. (1989). Differential treatment selection: The role of diagnosis in psychotherapy. *Psychotherapy*, 26, 271–281.
- Beutler, L. E. (2000). David and Goliath: When empirical and clinical standards of practice meet. *American Psychologist*, 55, 997–1007.
- Beutler, L. E. (2011). Prescriptive matching and systematic treatment selection. In J. C. Norcross, G. R. VandenBos, & D. K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (2nd ed., pp. 402–407). Washington, DC: American Psychological Association.
- Beutler, L. E., Albanese, A. L., Fisher, D., Karno, M., Sandowicz, M., Williams, O. B., ... & Thompson, L. (1999). *Selecting and matching to patient variables*. Paper presented at the annual meeting of the Society for Psychotherapy Research, Braga, Portugal.
- Beutler, L. E., & Clarkin, I. F. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions*. New York, NY: Brunner/Mazel.
- Beutler, L. E., Clarkin, I. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient*. New York, NY: Oxford University Press.
- Beutler, L. E., & Crago, M. (1986). Strategies and techniques of prescriptive psychotherapeutic intervention. In R. E. Hales & A. J. Francis (Eds.), *American Psychiatric Association Annual Review* (Vol. 6, pp. 378–397). Washington, DC: American Psychiatric Press.

- Beutler, L. E., Crago, M., & Arizmendi, T. G. (1986). Therapist variables in psychotherapy process and outcome. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavioral change* (3rd ed., pp. 257–310). New York, NY: Wiley.
- Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology*, 59, 1–8.
- Beutler, L. E., Forrester, B., Gallagher-Thompson, D., Thompson, L., & Tomlins, J. B. (2012). Common, specific, and treatment fit variables in psychotherapy outcome. *Journal of Psychotherapy Integration*, 22(3), 255–281.
- Beutler, L. E., Forrester, B., Holt, H., & Stein, M. (2013). Common, specific, and cross-cutting psychotherapy interventions. *Psychotherapy*, 50(3), 298–301.
- Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy*. New York, NY: Oxford University Press.
- Beutler, L. E., Harwood, T. M., Alimohamed, S., & Malik, M. (2002). Functional impairment and coping style. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs* (pp. 145–170). New York, NY: Oxford University Press.
- Beutler, L. E., Harwood, T. M., Bertoni, M., & Thomann, J. (2006). Systematic treatment selection and prescriptive therapy. In G. Striker & J. Gold (Eds.), *A casebook of psychotherapy integration*. Washington, DC: American Psychological Association Press.
- Beutler, L. E., Harwood, T. M., Kimpara, S., Verdirame, D., & Blau, K. (2011). Coping style. *Journal of Clinical Psychology*, 67(2), 176–183.
- Beutler, L. E., Harwood, T. M., Michelson, A., Song, X., & Holman, J. (2011). Resistance/reactance level. *Journal of Clinical Psychology*, 67(2), 133–142.
- Beutler, L. E., & Malik, M. (Eds.). (2002). *Rethinking the DSM: A psychological perspective*. Washington, DC: American Psychological Association.
- Beutler, L. E., Malik, M., Talebi, H., Fleming, J., & Moleiro, C. (2004). Use of psychological tests/instruments for treatment planning. In M. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed., pp. 101–132). Mahwah, NJ: Erlbaum.
- Beutler, L. E., & Moleiro, C. (2001). Clinical versus reliable and significant change. *Clinical Psychology: Science and Practice*, 8, 441–445.
- Beutler, L. E., Moleiro, C., Malik, M., & Harwood, T. M. (2003). A new twist on empirically supported treatments. *Internal Journal of Clinical and Health Psychology*, 3, 423–437.
- Beutler, L. E., Moleiro, C., & Talebi, H. (2002). Resistance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs* (pp. 129–444). New York, NY: Oxford University Press.
- Beutler, L. E., Sandowicz, M., Fisher, D., & Albanese, A. L. (1996). Resistance in psychotherapy: What can be concluded from empirical research. In *Session: Psychotherapy in Practice*, 2, 77–86.
- Beutler, L. E., & Williams, O. B. (1999). *Systematic treatment selection: A software package for treatment planning*. Ventura, CA: Center for Behavioral Technology.
- Beutler, L. E., Williams, R. E., & Norcross, J. C. (2008). *Innerlife.com: A software package for treatment planning*. Available at www.innerlife.com and www.webpsychcorp.com.
- Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State-Trait Anxiety Inventory, Trait version: Structure and content re-examined. *Behaviour Research and Therapy*, 36, 777–788.
- Bienvenu, O. J., Samuels, J. F., Riddle, M. A., Hoehn-Saric, R., Liang, K., Cullen, B. A. M., ... Nestadt, G. (2000). The relationship of obsessive-compulsive disorder to possible spectrum disorders: Results from a family study. *Biological Psychiatry*, 48, 287–293.
- Birchler, G. R. (1989). Review of *Behavioral assessment: A practical handbook* (3rd ed.). *Behavioral Assessment*, 11, 384–388.
- Birchler, G. R., & Fals-Stewart, W. (2006). Marital dysfunction. In M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 297–324). New York, NY: Elsevier.

- Black, F. W., & Bernard, B. A. (1984). Constructional apraxia as a function of lesion locus and size in patients with focal brain damage. *Cortex*, 20, 111–120.
- Black, J. (2000). Personality testing and police selection: Utility of the “Big Five.” *New Zealand Journal of Psychology*, 29(1), 2–9.
- Blais, M. A., Benedict, K. B., & Norman, D. K. (1995). Concurrent validity of the MCMI-II modifier indices. *Journal of Clinical Psychology*, 51(6), 783–789.
- Blalock, L. D., & McCabe, D. P. (2011). Proactive interference and practice effects in visuospatial working memory span task performance. *Memory*, 19(1), 83–91.
- Blanchard, E. B., Schwarz, S. P., Neff, D. F., & Gerardi, M. A. (1988). Prediction of outcome from the self-regulatory treatment of irritable bowel syndrome. *Behavior Research and Therapy*, 26, 187–190.
- Blatt, S. J. (2004). *Experiences of depression: Theoretical, clinical, and research perspectives*. Washington, DC: American Psychological Association.
- Blonigen, D. M., Timko, C., Jacob, T., & Moos, R. H. (2015). Patient-centered feedback on the results of personality testing increases early engagement in residential substance use disorder treatment: A pilot randomized controlled trial. *Addiction Science & Clinical Practice*, 10, 9.
- Blount, A., Schoenbaum, M., Kathol, R., Rollman, B. L., Marshall, T., O'Donohue, W., & Peek, C. J. (2007). The economics of behavioral health services in medical settings: A summary of evidence. *Professional Psychology: Research and Practice*, 38, 290–297.
- Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice*, 30, 252–259.
- Bockian, N., Meagher, S., & Millon, T. (2000). Assessing personality with the Millon Behavioral Health Inventory, the Millon Behavioral Medicine Diagnostic, and the Millon Clinical Multiaxial Inventory. In R. J. Gatchel & J. N. Weisberg (Eds.), *Personality characteristics of patients with pain* (pp. 61–88). Washington, DC: American Psychological Association.
- Boerger, A. R. (1975). *The utility of some alternative approaches to MMPI scale construction*. Unpublished doctoral dissertation, Kent State University, Kent, OH.
- Bolinsky, P. K., Trumbetta, S. L., Hanson, D. R., & Gottesman, I. I. (2010). Predicting adult psychopathology from adolescent MMPIs: Some victories. *Personality and Individual Differences*, 49(4), 324–330.
- Boll, T. J. (1974). Behavioral correlates of cerebral damage in children age 9–14. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and application*. Washington, DC: Winston & Sons.
- Bonarius, H. (1984). Prediction or anticipation: Some implications of personal construct psychology for professional practice. *Jyvaskyla Studies in Education: Psychology and Social Research*, 54, 190–206.
- Boone, D. (1998). Internal consistency reliability of the Personality Assessment Inventory with psychiatric inpatients. *Journal of Clinical Psychology*, 54(6), 839–843.
- Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY: Guilford Press.
- Bornstein, R. A. (1983). Verbal I.Q.–Performance I.Q. discrepancies on the Wechsler Adult Intelligence Scale–Revised in patients with unilateral or bilateral cerebral dysfunction. *Journal of Consulting and Clinical Psychology*, 51, 779–789.
- Bornstein, R. A., & Chelune, G. J. (1988). Factor analysis of the Wechsler Memory Scale–Revised. *Clinical Neuropsychologist*, 2, 107–115.
- Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency Scale: 1967–1995. *Psychological Assessment*, 8(2), 200–205.
- Bornstein, R. F. (1998). Reconceptualizing personality disorder diagnosis in DSM-V: The discriminant validity challenge. *Clinical Psychology: Science and Practice*, 5, 333–343.
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, 11, 48–57.

- Bornstein, R. F., & Masling, J. M. (2005). The Rorschach Oral Dependency Scale. In R. F. Bornstein & J. M. Masling (Eds.), *Scoring the Rorschach: Seven validated systems* (pp. 114–134). Mahwah, NJ: Erlbaum.
- Borus, J. F., Howes, M. J., Devins, N. P., & Rosenberg, R. (1988). Primary health care providers' recognition and diagnosis of mental disorders in their patients. *General Hospital Psychiatry*, 10, 317–321.
- Bowler, R. M., Thakler, C. D., & Becker, C. E. (1986). California Neuropsychological Screening Battery (CNC/BI & II). *Journal of Clinical Psychology*, 42, 946–955.
- Bracken, B. A., & McCallum, R. S. (1998). *The Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.
- Bracken, B. A., & McCallum, R. S. (2015). *Examiner's manual: Universal Nonverbal Intelligence Test, Second Edition (UNIT 2)*. Itasca, IL: Riverside.
- Brannigan, G., & Brunner, N. A. (1989). *The modified version of the Bender-Gestalt Test for Preschool and Primary School Children*. Branon, VT: Clinical Psychology Publishing.
- Brannigan, G., & Decker, S. L. (2003). *Bender Visual-Motor Gestalt Test II examiner's manual*. Itasca, IL: Riverside.
- Brannigan, G., & Decker, S. L. (2006). The Bender-Gestalt II. *American Journal of Orthopsychiatry*, 76(1), 10–12.
- Brawman-Mintzer, O., Lydiard, R. B., Emmanuel, N., Payeur, R., Johnson, M., Roberts, J., ... & Bellenger, J. C. (1993). Psychiatric comorbidity in patients with generalized anxiety disorder. *American Journal of Psychiatry*, 150(8), 1216–1218.
- Brayfield, A. H. (Ed.). (1965). Testing and public policy. *American Psychologist*, 20, 857–1005.
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. New York, NY: Academic Press.
- Breitholtz, E., Johansson, B., & Ost, L. G. (1999). Cognitions in generalized anxiety disorder and panic disorder patients: A prospective approach. *Behavior Research and Therapy*, 37, 533–544.
- Brenner, E. (2003). Consumer-focused psychological assessment. *Professional Psychology: Research and Practice*, 34, 240–247.
- Brewin, C. R. (1996). Theoretical foundations of cognitive-behavior therapy for anxiety and depression. *Annual Review of Psychology*, 47, 33–57.
- Bricklin, H. (1984). *Bricklin Perceptual Scales*. Furlong, PA: Village.
- Brod, M., Johnston, J., Able, S., & Swindle, R. (2006). Validation of the Adult Attention-deficit/Hyperactivity Disorder Quality-of-Life scale (AAQoL): A disease-specific quality-of-life measure. *Quality of Life Research*, 15(1), 117–129.
- Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34(10), 844–850.
- Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). The potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale–III in healthy older adults. *Journal of the International Neuropsychological Society*, 14, 463–478.
- Broshek, D. K., & Barth, J. T. (2000). The Halstead-Reitan Neuropsychological Test Battery. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 457–532). New York, NY: Wiley.
- Broughton, R. (1984). A prototype strategy for construction of personality scales. *Journal of Personality and Social Psychology*, 47, 1334–1346.
- Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). On the factor structure of the Beck Depression Inventory–II: G is the key. *Psychological Assessment*, 25(1), 136–145.
- Brown, L. (1990). Taking account of gender in the clinical assessment interview. *Professional Psychology*, 21, 12–17.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence-4 (TONI-4)*. Austin, TX: PRO-ED.

- Brown, T. A., O'Leary, T. A., & Barlow, D. H. (1993). Generalized anxiety disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (2nd ed., pp. 127–188). New York, NY: Guilford Press.
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology*, 87, 185–191.
- Bufka, L. F., Crawford, J. I., & Levitt, J. (2002). Brief screening instruments for managed care and primary care. In M. M. Antony & D. Barlow (Eds.), *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders* (pp. 49–62). New York, NY: Guilford Press.
- Bühler, J., Keller, F., and Läge, D. (2014). Activation as an overlooked factor in the BDI-II: A factor model based on core symptoms and qualitative aspects of depression. *Psychological Assessment*, 26(3), 970–979.
- Buhler, C., & LeFever, D. (1947). A Rorschach study on the psychological characteristics of alcoholics. *Quarterly Journal of Studies on Alcoholism*, 8, 197–260.
- Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin*, 117, 285–305.
- Butcher, J. N. (1990). *The MMPI-2 in psychological treatment*. New York, NY: Oxford University Press.
- Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press.
- Butcher, J. N. (1998). *Butcher Treatment Planning Inventory*. San Antonio, TX: Psychological Corporation.
- Butcher, J. N. (2004). Personality assessment without borders: Adaptation of the MMPI-2 across cultures. *Journal of Personality Assessment*, 83, 90–104.
- Butcher, J. N. (2006). *MMPI-2: A practitioner's guide* (2nd ed.). Washington, DC: American Psychological Association.
- Butcher, J. N. (2011). *A beginner's guide to the MMPI-2* (3rd ed.). Washington, DC: American Psychological Association.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring: MMPI-2*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring* (2nd ed.). Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 25–50). Hillsdale, NJ: Erlbaum.
- Butcher, J. N., & Perry, J. N. (2008). *Personality assessment in treatment planning: Use of the MMPI-2 and BTPI*. New York, NY: Oxford University Press.
- Butcher, J. N., Perry, J. N., & Hahn, J. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology*, 60, 331–345.
- Butcher, J. N., & Pope, K. S. (1989). The research base, psychometric properties, and clinical uses of the MMPI-2 and MMPI-A. *Canadian Psychology*, 33(1), 61–78.
- Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A interpretation* (2nd ed.). Minneapolis: University of Minnesota Press.
- Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., et al. (1992). *MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

- Butters, N., Grant, I., Haxby, J., Judd, L. L., Martin, A., McClelland, J., ... Stover, E. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH Workgroup on neuropsychological assessment approaches. *Journal of Clinical and Experimental Neuropsychology*, 12, 963–978.
- Cabral, R. R., & Smith, T. B. (2011). Racial/ethnic matching of clients and therapists in mental health services: A meta-analytic review of preferences, perceptions, and outcomes. *Journal of Counseling Psychology*, 58(4), 537–554.
- Cacciola, J. S., Alterman, A. I., Rutherford, M. J., McKay, J. R., & May, D. J. (1999). Comparability of telephone and in-person structured clinical interview for DSM-III-R (SCID) diagnoses. *Assessment*, 6, 235–242.
- Cacioppo, J. T., Glass, C. R., & Merluzzi, T. V. (1979). Self-statements and self-evaluations: A cognitive response analysis of heterosocial anxiety. *Cognitive Therapy and Research*, 3, 249–262.
- Caldwell, A. B. (1988). *MMPI Supplemental Scale manual*. Los Angeles, CA: Caldwell Report.
- Caldwell, A. B. (2001). What do the MMPI scales fundamentally measure? Some hypotheses. *Journal of Personality Assessment*, 76, 1–17.
- Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. R. (2000). Effects of response sets on NEO-PI-R scores and their relations to external criteria. *Journal of Personality Assessment*, 74(3), 472–488.
- Callaghan, G. M. (2001). Demonstrating clinical effectiveness for individual practitioners and clinics. *Professional Psychology: Research and Practice*, 32, 289–297.
- Calvin, J. (1975). *A replicated study of the concurrent validity of the Harris subscales for the MMPI*. Unpublished doctoral dissertation, Kent State University, Kent, OH.
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, 31, 141–154.
- Campbell, K. A., Rohlman, D. S., Storybach, D., Binder, L., Anger, W. K., Kovera, C. A., ... Grossman, S. J. (1999). Test-retest reliability of psychological and neurobehavioral tests set administered by computer. *Assessment*, 6, 21–32.
- Campos, R. C., & Gonçalves, B. (2011). The Portuguese version of the Beck Depression Inventory–II (BDI–II): Preliminary psychometric data with two nonclinical samples. *European Journal of Psychological Assessment*, 27(4), 258–264.
- Canetti, L., Shalev, A. Y., & De-Nour, A. K. (1994). Israeli adolescent norms of the Brief Symptom Inventory (BSI). *Israel Journal of Psychiatry and Related Sciences*, 31, 13–18.
- Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*, 29, 38–51.
- Canivez, G. L., & Watkins, M. W. (in press). Review of the Wechsler Intelligence Scale for Children–Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson, *Intelligence testing with the WISC-V*. Hoboken, NJ: Wiley.
- Caparrós-Caparrós, B., Villar-Hoz, E., Juan-Ferrer, J., & Viñas-Poch, F. (2007). Symptom Check-List–90–R: Fiabilidad, datos normativos y estructura factorial en estudiantes universitarios. *International Journal of Clinical and Health Psychology*, 7(3), 781–794.
- Carkhuff, R. R. (1969). *Helping and human relations. I: Selection and training. II: Practice and research*. New York, NY: Holt, Rinehart and Winston.
- Carlozzi, N. E., Horner, M. D., Yang, C., & Tilley, B. C. (2008). Factor analysis of the Repeatable Battery for the Assessment of Neuropsychological Status. *Applied Neuropsychology*, 15(4), 274–279.
- Carlson, J. F., Geisinger, K. F., & Jonson, J. L. (Eds.) (2014). *The nineteenth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.
- Carroll, L. M., Jaibert, A., Penney, A. M., Neath, I., Suprenant, A. M., & Tehan, G. (2010). Evidence for proactive interference in the focus of attention of working memory. *Canadian Journal of Experimental Psychology*, 64(3), 208–214.
- Cartwright, R. D. (1986). Affect and dream work from an information processing point of view. *Journal of Mind and Behavior*, 7, 411–428.

- Casacalenda, N., Perry, J. C., & Looper, K. (2002). Remission in major depressive disorder: A comparison of pharmacotherapy, psychotherapy, and control conditions. *American Journal of Psychiatry*, 159(8), 1354–1360.
- Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology*, 33, 446–453.
- Cashel, M. L., Rogers, R., Sewell, K., & Martin-Cannici, C. (1995). The Personality Assessment Inventory and the detection of defensiveness. *Assessment*, 2, 333–342.
- Castonguay, L. G., & Beutler, L. E. (2006). Common and unique principles of therapeutic change: What do we know and what do we need to know? In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work*. (pp. 353–370). New York, NY: Oxford University Press.
- Cella, D. (1984). The modified WAIS-R: An extension and revision. *Journal of Clinical Psychology*, 40, 801–804.
- Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P. J., Tabrizi, M. A., & Davies, M. (1985). The assessment of affective disorders in children and adolescents by semistructured interview: Test-retest reliability of the Schedule for Affective Disorders and Schizophrenia for School Age Children, Present Episode version. *Archives of General Psychiatry*, 42, 696–702.
- Chang, H. (2005). Dimensions of the Chinese Beck Depression Inventory–II in a university sample. *Individual Differences Research*, 3(3), 193–199.
- Charter, R. A., & Lopez, M. N. (2002). Millon Clinical Multiaxial Inventory (MCMI–II/MCMI–III): The inability of the validity conditions to detect random responders. *Journal of Clinical Psychology*, 58, 1615–1617.
- Chelune, G. J., & Bornstein, R. A. (1988). WMS-R patterns among patients with unilateral brain lesions [Special issue]. *Clinical Neuropsychologist*, 2, 121–132.
- Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, 49(7), 677–682.
- Cheung, F. M., Cheung, S. F., Leung, K., Ward, C., & Leong, F. (2003). The English version of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology*, 34(4), 433–452.
- Cheung, F. M., & Ho, R. M. (1997). Standardization of the Chinese MMPI-A in Hong Kong: A preliminary study. *Psychological Assessment*, 9(4), 499–502.
- Cheung, F. M., Leung, K., Fan, R. M., Song, W., Zhang, J., & Zhang, J. (1996). *Journal of Cross-Cultural Psychology*, 27(2), 181–199.
- Cheung, F. M., Leung, K., Zhang, J., Sun, H., Gan, Y., Song, W., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the five-factor model complete? *Journal of Cross-Cultural Psychology*, 32(4), 407–433.
- Chiles, J. A., Lambert, M. J., & Hatch, A. L. (1999). The impact of psychological interventions on medical cost offset: A meta-analytic review. *Clinical Psychology: Science and Practice*, 6, 204–220.
- Chisholm, S. M., Crowther, J. H., & Ben-Porath, Y. S. (1997). Selected MMPI-2 Scales' ability to predict premature termination and outcome from psychotherapy. *Journal of Personality Assessment*, 69, 127–144.
- Choca, J. P., Retzlaff, P., Strack, S., Mouton, A., & Van Denburg, E. (1996). Factorial elements in Millon's personality theory. *Journal of Personality Disorders*, 10, 377–383.
- Choca, J. P., Shanley, L. A., Peterson, C. A., & Van Denburg, E. (1990). Racial bias and the MCMI. *Journal of Personality Assessment*, 54, 479–490.
- Choca, J. P., & Van Denburg, E. (2004). *Interpretive guide to the Millon Clinical Multiaxial Inventory* (3rd ed.). Washington, DC: American Psychological Association.
- Choca, J. P., Van Denburg, E., Bratu, M. E., & Meagher, S. (1995, March). *Personality changes of psychiatric patients with aging*. Paper presented at the midwinter meeting of the Society for Personality Assessment, Denver, CO.

- Chojnacki, J. T., & Walsh, W. B. (1992). The consistency of scores and configural patterns between MMPI and MMPI-2. *Journal of Personality Assessment*, 59, 276–289.
- Choquette, K. A. (1990). Assessing depression in alcoholics with the BDI, SCL-90-R, and DIS criteria. *Journal of Substance Abuse*, 6, 295–304.
- Chor, K. H. B., McClelland, G. M., Weiner, D. A., Jordan, N., & Lyons, J. S. (2012). Predicting outcomes of children in residential treatment: A comparison of a decision support algorithm and a multidisciplinary team decision model. *Children and Youth Services Review*, 34(12), 2345–2352.
- Chor, K. H. B., McClelland, G. M., Weiner, D. A., Jordan, N., & Lyons, J. S. (2013). Patterns of out-of-home placement decision-making in child welfare. *Child Abuse & Neglect*, 37(10), 871–882.
- Chor, K. H. B., McClelland, G. M., Weiner, D. A., Jordan, N., & Lyons, J. S. (2015). Out-of-home placement decision-making and outcomes in child welfare: A longitudinal study. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(1), 70–86.
- Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., ... Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, 18, 153–172.
- Ciarrochi, J. V., Chan, A. Y. C., & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences*, 28(3), 539–561.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Ciminero, A. R., Calhoun, K. S., & Adams, H. E. (Eds.). (1977). *Handbook of behavioral assessment*. New York, NY: Wiley.
- Claassen, C. A., & Lovitt, R. (2001). Solving ethical problems in medical setting during psychological assessment: A decisional model. *Journal of Personality Assessment*, 77, 214–230.
- Clare, I. C. H., & Gudjonsson, G. H. (1995). The vulnerability of suspects with intellectual disabilities during police interviews: A review and experimental study of decision-making. *Mental Handicap Research*, 8, 110–128.
- Clark, C. G., & Miller, H. L. (1971). Validation of Gilberstadt and Duker's 8–6 profile type on a black sample. *Psychological Reports*, 29, 259–264.
- Clark, M. E. (1996). MMPI-2 negative treatment indicators content and content component scales: Clinical correlates and outcome prediction for men with chronic pain. *Psychological Assessment*, 8(1), 32–38.
- Coan, R. (1956). A factor analysis of Rorschach determinants. *Journal of Projective Techniques*, 20, 280–287.
- Cofer, C. N., Chance, J., & Judson, A. J. (1949). A study of malingering on the Minnesota Multiphasic Personality Inventory. *Journal of Psychology: Interdisciplinary and Applied*, 27(2), 491–499.
- Colby, F. (1989). Usefulness of the K correction in MMPI profiles of patients and non-patients. *Psychological Assessment*, 1, 142–145.
- Colligan, R. C., & Offord, K. P. (1989). The aging MMPI: Contemporary norms for contemporary teenagers. *Mayo Clinic Proceedings*, 64, 3–27.
- Cone, J. D. (1977). The relevance of reliability and validity for behavioral assessment. *Behavior Therapy*, 8, 411–426.
- Cone, J. D. (1978). The behavioral assessment grid (BAG): A conceptual framework and taxonomy. *Behavior Therapy*, 9, 882–888.
- Cone, J. D. (1998). Psychometric considerations: Concepts, contents, and methods. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (4th ed., pp. 22–46). Boston, MA: Allyn & Bacon.
- Conners, C. K. (2008). *Conners-3 Parent and Teacher Rating Scales*. North Tonawanda, NY: Multi-Health Systems.

- Conners, C. K. (2015). *Conners CPT 3: Conners Continuous Performance Test, 3rd Edition*. North Tonawanda, NY: Multi-Health Systems.
- Constantino, M. J., Glass, C. R., Arnkoff, D. B., Ametrano, R. M., & Smith, J. Z. (2011). Expectations. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 354–376). New York, NY: Oxford University Press.
- Cooney, N. L., Kadden, R. M., & Litt, M. D. (1990). A comparison of methods assessing sociopathy in male and female alcoholics. *Journal of Studies on Alcohol*, 51, 42–48.
- Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87, 442–449.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218–244.
- Cooper, Z., & Fairburn, C. G. (1987). The eating disorder examination: A semistructured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders*, 6, 136–141.
- Corbella, S., Beutler, L. E., Fernandez-Alvarez, H., Botella, L., Malik, M. L., Lane, G., & Wagstaff, N. (2003). Measuring coping style and resistance among Spanish and Argentine samples: Development of the Systematic Treatment Selection Self-Report in Spanish. *Journal of Clinical Psychology*, 59(9), 921–932.
- Corkin, S., Growdon, J. H., Sullivan, E. V., Nissen, M. J., & Huff, F. J. (1986). Assessing treatment effects: A neuropsychological battery. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults* (pp. 156–167). Washington, DC: American Psychological Association.
- Cormier, W. H., & Cormier, L. S. (1998). *Interviewing strategies for helpers: Fundamental skills and cognitive behavioral interventions* (4th ed.). Monterey, CA: Brooks/Cole.
- Corsica, J. A., Azarbad, L., McGill, K., Wool, L., & Hood, M. (2010). The Personality Assessment Inventory: Clinical utility, psychometric properties, and normative data for bariatric surgery candidates. *Obesity Surgery*, 20(6), 722–731.
- Coryell, W. H., Akiskal, H. S., Leon, A. C., Winokur, G., Masur, J. D., Mueller, T. I., & Keller, M. B. (1994). The time course of nonchronic major depressive disorder. *Archives of General Psychiatry*, 51, 405–410.
- Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory: Manual Form S and Form R*. Lutz, FL: Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (1990). Personality disorders and the five-factor model of personality. *Journal of Personality Disorders*, 4(4), 362–371.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO Personality Inventory. *Journal of Personality Assessment*, 64(1), 21–50.
- Costa, P. T., Jr., & McCrae, R. R. (2003). *Bibliography for the NEO-PI-R and NEO-FFI*. Lutz, FL: Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (2009). The five-factor model and the NEO inventories. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment*. New York, NY: Oxford University Press.
- Costa, P. T., Jr., & McCrae, R. R. (2014). The NEO Inventories. In R. P. Archer & S. R. Smith (Eds.), *Personality Assessment* (2nd ed., pp. 229–260). New York, NY: Routledge.
- Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, 12(9), 887–898.
- Costa, P. T., Jr., & Piedmont, R. L. (2005). Multivariate assessment: NEO-PI-R profiles of Madeline G. In J. S. Wiggins (Ed.), *Paradigms of personality assessment*. New York, NY: Guilford Press.
- Costa, P. T., Jr., Stone, S. V., McCrae, R. R., Dembroski, T. M., & Williams, R. B. (1987). Hostility, agreeableness-antagonism, and coronary heart disease. *Holistic Medicine*, 2, 161–167.

- Costa, P. T., Jr., Zonderman, A. B., & McCrae, R. R. (1991). Personality, defense, coping, and adaptation in older adulthood. In E. M. Cummings, A. L. Greene, & K. H. Karraker (Eds.), *Lifespan developmental psychology: Perspectives on stress and coping* (pp. 277–293). Hillsdale, NJ: Erlbaum.
- Costa, P. T., Jr., Zonderman, A. B., Williams, R. B., & McRae, R. R. (1985). Content and comprehensiveness of the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology*, 48, 925–933.
- Costello, E. J., Benjamin, R., Angold, A., & Silver, D. (1991). Mood variability in adolescence: A study of depressed and nondepressed comorbid patients. *Journal of Affective Disorders*, 23, 199–212.
- Costello, E. J., Edelbrock, C. S., & Costello, A. J. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal and Child Psychology*, 13, 579–595.
- Costello, E. J., Edelbrock, C. S., Duncan, M. K., & Kalas, R. (1984). *Testing of the NIMH Diagnostic Interview Schedule for Children (DISC) in a clinical population. Final report to the Center for Epidemiological Studies, National Institute for Mental Health*. Pittsburgh, PA: University of Pittsburgh.
- Craig, R. J. (1999). Overview and status of the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment*, 72, 390–406.
- Craig, R. J. (2001). Adjectival descriptions of personality disorders: A convergent validity study of the MCMI-IIMCMI-III. *Journal of Personality Assessment*, 77, 259–271.
- Craig, R. J. (2005). Alternative interpretations for the histrionic, narcissistic, and compulsive personality disorder scales of the MCMI-III. In R. J. Craig (Ed.), *New directions in interpreting the Millon Clinical Multiaxial Inventory-III* (pp. 71–93). Hoboken, NJ: Wiley.
- Craig, R. J., & Bivens, A. (1998). Factor structure of the MCMI-IIMCMI-III. *Journal of Personality Assessment*, 70, 190–196.
- Craig, R. J., & Olson, R. E. (2001). Adjectival descriptions of personality disorders: A convergent validity study of the MCMI-IIMCMI-III. *Journal of Personality Assessment*, 77, 259–271.
- Craighead, W. E., & Dunlop, B. W. (2014). Combination psychotherapy and antidepressant medication treatment for depression: For whom, when, and how. *Annual Review of Psychology*, 65, 267–300.
- Craighead, W. E., Johnson, B. N., Carey, S., & Dunlop, B. W. (2015). Psychosocial treatments for major depressive disorder. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (4th ed., pp. 381–408). New York, NY: Oxford University Press.
- Crary, W. G., & Johnson, C. W. (1981). Mental status examination. In C. W. Johnson, J. R. Snibbe, & L. A. Evans (Eds.), *Basic psychopathology: A programmed text* (2nd ed., pp. 56–57). Lancaster, PA: MTP Press.
- Craske, M. G., & Barlow, D. H. (1993). Panic and agoraphobia. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 1–47). New York, NY: Guilford Press.
- Crumpton, E. (1956). The influence of color on the Rorschach Test. *Journal of Projective Techniques*, 20, 150–158.
- Cruz Fuentes, C. S., Bellow, L. L., Garcia, C. B., Macías, L. G., & Chavez Balderas, R. A. (2005). Datos sobre la validez y confiabilidad de la Symptom Check List 90 (SCL 90) en una muestra de sujetos mexicanos. *Salud Mental*, 28(1), 72–81.
- Cucciare, M. A., & Weingardt, K. R. (Eds.). (2009). *Using technology to support evidence-based behavioral health practices: A clinician's guide*. New York, NY: Routledge.
- Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican Americans-II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Science*, 17, 275–304.
- Cummings, N. A. (1991). The somatizing patient. In C. A. Austad & W. A. Berman (Eds.), *Psychotherapy in managed care: The optimal use of time and resources* (pp. 234–237). Washington, DC: American Psychological Association.
- Cunningham, T. R., & Thorp, R. G. (1981). The influence of settings on accuracy and reliability of behavioral observation. *Behavioral Assessment*, 3, 67–78.

- Cyr, J. J., McKenna-Foley, J. M., & Peacock, E. (1985). Factor structure of the SCL-90-R: Is there one? *Journal of Personality Assessment*, 49, 571-578.
- Dahlstrom, W. G. (1969). Recurrent issues in the development of the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications* (pp. 1-40). New York, NY: McGraw-Hill.
- Dahlstrom, W. G., Lachar, D., & Dahlstrom, L. E. (1986). Overview and conclusions. In W. G. Dahlstrom, D. Lachar, & L. E. Dahlstrom (Eds.), *MMPI patterns of American minorities* (pp. 188-205). Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., & Welsh, G. S. (1960). *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook. Vol. 1: Clinical interpretation*. Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook. Vol. 2: Research developments and applications*. Minneapolis: University of Minnesota Press.
- Daldrup, F. J., Beutler, L. E., Engle, D., & Greenberg, L. S. (1988). *Focused expressive psychotherapy: Freeing the overcontrolled patient*. New York, NY: Pergamon Press.
- Dana, R. H. (1965). Review of the Rorschach. In O. K. Buros (Ed.), *Sixth mental measurements yearbook* (pp. 492-495). Highland Park, NJ: Gryphon Press.
- Dana, R. H. (1968). Six constructs to define Rorschach M. *Journal of Projective Techniques and Personality Assessment*, 32, 138-145.
- Dana, R. H. (2005). *Multicultural assessment: Principles, applications, and examples*. Mahwah, NJ: Erlbaum.
- Dana, R. H., Field, K., & Bolton, B. (1983). Variations of the Bender-Gestalt Test: Implications for training and practice. *Journal of Personality Assessment*, 47, 76-84.
- Dao, T. K., & Prevatt, F. (2006). A psychometric evaluation of the Rorschach System's Perceptual Thinking Index. *Journal of Personality Assessment*, 86, 180-189.
- Davis, A., Williams, R. N., Gupta, A. S., Finch, W. H., & Randolph, C. (2015). Evaluating neurocognitive deficits in patients with multiple sclerosis via a brief neuropsychological approach. *Applied Neuropsychology: Adult*, 22(5), 381-387.
- Davis, R. G. (1999). Millon: Essentials of his science, theory, classification, assessment, and therapy. *Journal of Personality Assessment*, 72, 330-352.
- Deardorff, W. W. (2000). The MMPI-2 and chronic pain. In R. J. Gatchel & J. N. Weisberg (Eds.), *Personality characteristics of patients with pain* (pp. 109-125). Washington, DC: American Psychological Association.
- Debord, J. B. (1989). Paradoxical interventions: A review of the recent literature. *Journal of Counseling and Development*, 67, 394-398.
- Decker, S. L. (2007). Measuring growth and decline in visual-motor processes with the Bender-Gestalt Second Edition. *Journal of Psychoeducational Assessment*, 10, 1-13.
- Decker, S. L., Allen, R., & Choca, J. P. (2006). Construct validity of the Bender-Gestalt II: Comparison with the Wechsler Intelligence Scale for Children-III. *Perceptual and Motor Skills*, 102, 133-141.
- De Fruyt, F., McCrae, R. R., Szirmák, Z., & Nagy, J. (2004). The Five-Factor Personality Inventory as a measure of the five-factor model: Belgian, American, and Hungarian comparisons with the NEO-PI-R. *Assessment*, 11(3), 207-215.
- De Fruyt, F., & Mervielde, I. (1997). The five-factor model of personality and Holland's RIASEC interest types. *Personality and Individual Differences*, 23(1), 87-103.
- Deisinger, J. A. (1995). Exploring the factor structure of the Personality Assessment Inventory. *Assessment*, 2(2), 173-179.
- DeKoninck, J. M., & Crabbe-Decleve, G. (1971) Field dependence and Rorschach white-space figure-ground reversal responses. *Perceptual and Motor Skills*, 33(2), 1191-1194.
- de la Rubia, J. M. (2013). Validación de un format simplificado del Inventario de Depresión de Beck (BDI-2). *Psicología Iberoamericana*, 21(1), 42-52.
- Del Greco, L., Breitbach, L., & McCarthy, R. H. (1981). The Rathus Assertiveness Schedule modified for early adolescents. *Journal of Behavioral Assessment*, 3, 321-328.

- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System*. San Antonio, TX: Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test—II: Adult version*. San Antonio, TX: Psychological Corporation.
- den Boer, P. C., Wiersma, D., & van den Bosch, R. J. (2004). Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. *Psychological Medicine*, 34(6), 959–971.
- Derogatis, L. R. (1992). *BSI: Administration, scoring, and procedures manual—II* (2nd ed.). Baltimore, MD: Clinical Psychometric Research.
- Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI) administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.
- Derogatis, L. R. (1994). *SCL-90-R: Administration, scoring, and procedures manual*. Minneapolis, MN: National Computer Systems.
- Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCl): A self-report symptom inventory. *Behavioral Scientist*, 19, 1–15.
- Derogatis, L. R., & Savitz, K. L. (1999). The SCL-90-R, Brief Symptom Inventory, and matching clinical rating scales. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 679–724). Mahwah, NJ: Erlbaum.
- desRosiers, G. (1992). Primary or depressive dementia: Clinical features. *International Journal of Geriatric Psychiatry*, 7, 629–638.
- Dies, R. R. (1972). Personal gullibility or pseudodiagnosis: A further test of the “fallacy of personal validation.” *Journal of Clinical Psychology*, 28, 47–50.
- Dikmen, S. S., Corrigan, J. D., Levin, H. S., Machamer, J., Stiers, W., & Weisskopf, M. G. (2009). Cognitive outcome following traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 24(6), 430–438.
- DiNardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV*. Albany, NY: Graywind.
- Dobbins, C., & Russell, E. W. (1990). Left temporal lobe brain damage pattern on the Wechsler Adult Intelligence Scale. *Journal of Clinical Psychology*, 46, 863–868.
- Dolle, K., Schulte-Körne, G., O’Leary, A. M., von Hofacker, N., Izat, Y., & Allgaier, A. K. (2012). The Beck Depression Inventory—II in adolescent mental health patients: Cut-off scores for detecting depression and rating severity. *Psychiatry Research*, 200(2), 843–848.
- Dolman, R., Roy, E. A., Dimeck, P. T., & Hall, C. R. (2000). Age, gesture span, and dissociation among component subsystems of working memory. *Brain, Cognition, and Rehabilitation*, 43, 164–168.
- Donaldson, J. (2010). *The Personality Assessment Inventory, women, and poverty: Psychometric properties and clinical utility* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.
- Dong, Y. T., & Church, A. T. (2003). Cross-cultural equivalence and validity of the Vietnamese MMPI-2: Assessing psychological adjustment of Vietnamese refugees. *Psychological Assessment*, 15(3), 370–377.
- Donias, S., Karastergiou, A., & Manos, N. (1991). Standardization of the Symptom Checklist-90-R rating scale in a Greek population. *Psychiatriki*, 2(1), 42–48.
- Dorr, D. (1999). Approaching psychotherapy of the personality disorders from the Millon perspective. *Journal of Personality Assessment*, 72, 407–425.
- Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology*, 71, 9–15.
- Dowd, E. T., Milne, C. R., & Wise, S. L. (1991). The Therapeutic Reactance Scale: A measure of psychological reactance. *Journal of Counseling and Development*, 69, 541–545.
- Dowd, E. T., & Wallbrown, F. (1993). Motivational components of client reactance. *Journal of Counseling and Development*, 71, 533–538.
- Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory—II. *Psychological Assessment*, 10, 83–89.

- Draguns, J. G., Haley, E. M., & Phillips, L. (1967). Studies of the Rorschach content: A review of the research literature. Part 1: Traditional content categories. *Journal of Projective Techniques and Personality Assessment*, 31, 3–32.
- Dritschel, B. H., Williams, K., & Cooper, P. J. (1991). Cognitive distortions among women experiencing bulimic episodes. *International Journal of Eating Disorders*, 10, 547–555.
- Drozdzick, L. W., & Collum, C. M. (2011). Expanding the ecological validity of WAIS-IV and WMS-IV with the Texas Functional Living Scale. *Assessment*, 18(2), 141–155.
- Drozdzick, L. W., Holdnack, J. A., & Hilsabeck, R. C. (2011). *Essentials of WMS-IV assessment*. Hoboken, NJ: Wiley.
- DuAlba, L., & Scott, R. L. (1993). Somatization and malingering for workers' compensation applicants: A cross-cultural MMPI study. *Journal of Clinical Psychology*, 49, 913–917.
- Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, O., Cummings, J., ... & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurology*, 6, 734–746.
- Dudek, S. Z. (1968). Regression and creativity: A comparison of the Rorschach records of successful vs. unsuccessful painters and writers. *Journal of Nervous and Mental Disease*, 147(6), 535–546.
- Duff, K., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., Mold, J., ... Adams, R. L. (2006). Examining the Repeatable Battery for the Assessment of Neuropsychological Status: Factor analytic studies in an elderly sample. *American Journal of Geriatric Psychiatry*, 14, 976–979.
- Duff, K., Leigh, J. B., Schoenberg, M. R., Patton, D. E., Mold, J., Scott, J. G., & Adams, R. (2005). Test-retest stability and practice effects of the RBANS in a community-dwelling elderly sample. *Journal of Clinical and Experimental Neuropsychology*, 27, 565–575.
- Duff, K., Patton, D., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. (2003). Age and education-corrected independent normative data for the RBANS in a community-dwelling elderly sample. *Clinical Neuropsychologist*, 17, 351–366.
- Dumitrascu, N., Mihura, J. L., & Meyer, G. J. (2011, March). *Selected R-PAS variables and their relationship to level of education (as proxy for cognitive sophistication)*. Paper presented at the annual meeting of the Society for Personality Assessment, Boston, MA.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Minneapolis, MN: Pearson.
- Durand, V. M., & Crimmins, D. B. (1992). *The Motivation Assessment Scale administrative guide*. Topeka, KS: Monaco & Associates.
- Duthie, B., & Vincent, K. R. (1986). Diagnostic hit rates of high point codes for the Diagnostic Inventory of Personality and Symptoms using random assignment, base rates, and probability scales. *Journal of Clinical Psychology*, 42, 612–614.
- Earls, R., Reich, W., Jung, K. G., & Cloninger, C. R. (1988). Psychopathology in children of alcoholic and antisocial parents. *Alcoholism: Clinical and Experimental Research*, 12, 481–487.
- Eaton, W. W., Neufeld, K., Chen, L., & Cai, G. (2000). A comparison of self report and clinical diagnostic interviews for depression. *Archives of General Psychiatry*, 57, 217–222.
- Edell, W. S. (1987). Role of structure in disordered thinking in borderline and schizophrenic disorders. *Journal of Personality Assessment*, 51, 23–41.
- Edens, J. F., Cruise, K. R., & Buffington-Vollum, J. K. (2001). Forensic and correctional applications of the Personality Assessment Inventory. *Behavioral Sciences and the Law*, 19, 519–543.
- Edwards, D. W., Morrison, T. L., & Weissman, H. N. (1993). The MMPI and MMPI-2 in an outpatient sample: Comparisons of code types, validity scales, and clinical scales. *Journal of Personality Assessment*, 61, 1–18.
- Eells, T. D. (2011). *Handbook of psychotherapy case formulation*. New York, NY: Guilford Press.
- Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of a content coding manual. *Journal of Psychotherapy Practice and Research*, 7(2), 144–153.

- Egeland, B. R. (1969). Examiner expectancy: Effects on the scoring of the WISC. *Psychology in the Schools*, 6, 313–315.
- Eich, D., Angst, J., Frei, A., Ajdacic-Gross, V., Rössler, W., & Gamma, A. (2012). A new rating scale for adult ADHD based on the Symptom Checklist 90 (SCL-90-R). *European Archives of Psychiatry and Clinical Neuroscience*, 262(6), 519–528.
- Eidelson, R. J., & Epstein, N. (1982). Cognitive and relationship adjustment: Development of a measure of dysfunctional relationship beliefs. *Journal of Consulting and Clinical Psychology*, 50, 715–720.
- Eisman, E. J., Dies, R. R., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W., ... Moreland, K. L. (2000). Problems and limitations in using psychological assessment in the contemporary health care delivery system. *Professional Psychology: Research and Practice*, 31, 131–140.
- Elashoff, J., & Snow, R. E. (Eds.). (1971). *Pygmalion revisited*. Worthington, OH: C. A. Jones.
- Elion, V. H., & Megargee, E. I. (1975). Validity of the MMPI Pd Scale among black males. *Journal of Consulting and Clinical Psychology*, 43, 166–172.
- Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 63(5), 841–847.
- Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., ... Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program. *Archives of General Psychiatry*, 46, 971–982.
- Elliot, A. J., Miltenberger, R. G., Kaster-Bundgaard, J., & Lumley, V. (1996). A national survey of assessment and therapy techniques used by behavior therapists. *Cognitive and Behavioral Practice*, 3, 107–125.
- Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, 48(1), 43–49.
- Elliott, R., Fox, C. M., Svetlana, A., B., Stone, G. E., Gunderson, J., & Zhang, X. (2006). Deconstructing therapy outcome measurement with Rasch analysis of a measure of general clinical distress: The Symptom Checklist-90-Revised. *Psychological Assessment*, 18, 359–372.
- Elwood, R. W. (1991). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychology Review*, 2, 179–201.
- Embretson, S., Schneider, L. M., & Roth, D. L. (1986). Multiple processing strategies and the construct validity of verbal reasoning tests. *Journal of Educational Measurement*, 23, 13–32.
- Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, 35, 837–844.
- Eppinger, M. G., Craig, P. L., Adams, R. L., & Parsons, O. A. (1987). The WAIS-R Index for estimating premorbid intelligence: Cross-validation and clinical utility. *Journal of Consulting and Clinical Psychology*, 55, 86–90.
- Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in Human Behavior*, 17, 295–314.
- Equal Employment Opportunity Commission. (1970). Guidelines on employee selection procedures. *Federal Register*, 35, 12333–12336.
- Erginel, A. (1972). On the test-retest reliability of the Rorschach. *Journal of Personality Assessment*, 36, 203–212.
- Erickson, R. C., Caslyn, D. A., & Scheupbach, C. S. (1978). Abbreviating the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology*, 42, 946–955.
- Essau, C. A., Groen, G., Conradt, J., Turbanisch, U., & Petermann, F. (2001). Validität und Reliabilität der SCL-90-R: Ergebnisse der Bremer Jugendstudie. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22(2), 139–152.
- Exner, J. E. (1969). *The Rorschach systems*. New York, NY: Grune & Stratton.
- Exner, J. E. (1974). *The Rorschach: A comprehensive system. Volume 1: Basic foundations*. New York, NY: Wiley.
- Exner, J. E. (1978). *The Rorschach: A comprehensive system. Volume 2: Current research and advanced interpretations*. New York, NY: Wiley.

- Exner, J. E. (1983). Rorschach assessment. In I. B. Weiner (Ed.), *Clinical methods in clinical psychology* (2nd ed.). New York, NY: Wiley.
- Exner, J. E. (1984). *A computer program to assist in Rorschach interpretation* (rev. ed.). Bayville, NY: Rorschach Workshops.
- Exner, J. E. (1986). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (2nd ed.). New York, NY: Wiley.
- Exner, J. E. (1991). *The Rorschach: A comprehensive system. Volume 2: Current research and advanced interpretation* (2nd ed.). New York, NY: Wiley.
- Exner, J. E. (1993). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (3rd ed.). New York, NY: Wiley.
- Exner, J. E. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessing children and adolescents*. New York, NY: Wiley.
- Exner, J. E. (1997). The future of Rorschach in personality assessment. *Journal of Personality Assessment*, 68, 37–46.
- Exner, J. E. (2000). *A primer for Rorschach interpretation*. Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2001). *A Rorschach workbook for the comprehensive system* (5th ed.). Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2003). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (4th ed.). Hoboken, NJ: Wiley.
- Exner, J. E., Armbruster, G. L., & Viglione, D. (1978). The temporal stability of some Rorschach features. *Journal of Personality Assessment*, 42, 474–482.
- Exner, J. E., Armbruster, G. L., Walker, E. J., & Cooper, W. H. (1975). *Anticipation of elective surgery as manifest in Rorschach records*. Rorschach Workshops (Study No. 213, unpublished).
- Exner, J. E., & Bryant, E. L. (1974). *Rorschach responses of subjects recently divorced or separated*. Rorschach Workshops (Study No. 206, unpublished).
- Exner, J. E., & Erdberg, P. (2005). *The Rorschach: A comprehensive system, advanced interpretation*, (Vol. 2, 3rd ed.). Hoboken, NJ: Wiley.
- Exner, J. E., & Exner, D. E. (1972). How clinicians use the Rorschach. *Journal of Personality Assessment*, 36, 403–408.
- Exner, J. E., & Hillman, L. (1984). *A comparison of content distributions for the records of 76 paranoid schizophrenics and 76 nonparanoid schizophrenics*. Rorschach Workshops (Study No. 293, unpublished).
- Exner, J. E., & Kazaoka, K. (1978). *Dependency gestures of 16 assertiveness trainees as related to Rorschach movement responses*. Rorschach Workshops (Study No. 261, unpublished).
- Exner, J. E., Leura, A. V., & George, L. M. (1976). *A replication of the Masling study using four groups of new examiners with two seating arrangements and ride evaluation*. Rorschach Workshops (Study No. 256, unpublished).
- Exner, J. E., & Murillo, L. G. (1973). Effectiveness of regressive ECT with process schizophrenics. *Diseases of the Nervous System*, 34, 44–48.
- Exner, J. E., & Murillo, L. G. (1977). A long-term follow up of schizophrenics treated with regressive ECT. *Diseases of the Nervous System*, 38, 162–168.
- Exner, J. E., Murillo, L. G., & Sternklar, S. (1979). *Anatomy and X-ray responses among patients with body delusions or body problems*. Rorschach Workshops (Study No. 257, unpublished).
- Exner, J. E., & Sanglade, A. A. (1992). Rorschach changes following brief and short-term therapy. *Journal of Personality Assessment*, 59, 59–71.
- Exner, J. E., Thomas, E. A., Cohen, J. B., Ridgeway, E. M., & Cooper, W. H. (1981). *Stress indices in the Rorschachs of patients recovering from myocardial infarctions*. Rorschach Workshops (Study No. 286, unpublished).
- Exner, J. E., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Descriptions and prediction. *Journal of Personality Assessment*, 49, 13–20.
- Exner, J. E., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents*. New York, NY: Wiley.

- Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents* (2nd ed.). New York, NY: Wiley.
- Exner, J. E., & Wylie, J. R. (1975). *Attempts at simulation of schizophrenia-like protocols by psychology graduate students*. Rorschach Workshops (Study No. 211, unpublished). Bayville, NY: Rorschach Workshops.
- Exner, J. E., & Wylie, J. R. (1977). Some Rorschach data concerning suicide. *Journal of Personality Assessment*, 41, 339–348.
- Eysenck, H. J. (1994). The outcome problem in psychotherapy: What have we learned? *Behavior Research and Therapy*, 32, 477–495.
- Fallows, R. R., & Hilsabeck, R. C. (2012). WAIS-IV Visual Puzzles in a mixed clinical sample. *Clinical Neuropsychologist*, 26(6), 942–950.
- Fals-Stewart, W. (1995). The effect of defensive responding by substance-abusing patients on the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment*, 64, 540–541.
- Fals-Stewart, W. (1996). The ability of individuals with psychoactive substance use disorders to escape detection by the Personality Assessment Inventory. *Psychological Assessment*, 8(1), 60–68.
- Fantoni-Salvador, P., & Rogers, R. (1997). Spanish versions of the MMPI-2 and PAI: An investigation of concurrent validity with Hispanic patients. *Assessment*, 4(1), 29–39.
- Faust, D. (1991). Forensic neuropsychology: The art of practicing a science that does not yet exist. *Neuropsychology Review*, 2, 205–231.
- Faust, D., & Ziskin, J. (1989). Computer-assisted psychological evaluation as legal evidence: Someday my prints will come. *Computers in Human Behavior*, 5, 23–36.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26, 57–63.
- Feingold, A. (1983). The validity of the information and vocabulary subtests of the WAIS for predicting college achievement. *Educational and Psychological Measurement*, 43, 1127–1131.
- Feldman, S. E., & Sullivan, D. S. (1971). Factors mediating the efforts of enhanced rapport on children's performance. *Journal of Consulting and Clinical Psychology*, 36, 302.
- Fernandez, K., Boccaccini, M. T., & Noland, R. M. (2008). Detecting over- and underreporting of psychopathology with the Spanish-language Personality Assessment Inventory: Findings from a simulation study with bilingual speakers. *Psychological Assessment*, 20(2), 189–194.
- Filskov, S. B. (1978). *The prediction of impairment from figure copying*. Paper presented at the Southeastern Psychological Association convention, Atlanta, GA.
- Fincham, F. D., & Bradbury, T. N. (1992). Assessing attributions in marriage: The Relationship Attribution Measure. *Journal of Personality and Social Psychology*, 62, 457–468.
- Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, 11, 58–66.
- Fink, J., McCrea, M., & Randolph, C. (1998). Neuropsychological differentiation of vascular dementia and Alzheimer's disease: A neurocognitive profile approach using a short battery. *Journal of the International Neuropsychological Society*, 4, 30.
- Finn, S. E. (1982). Base rates, utilities, and the DSM-III: Shortcomings of fixed rule systems of psychodiagnostics. *Journal of Abnormal Psychology*, 48, 294–302.
- Finn, S. E. (1996). *A manual for using the MMPI-2 as a therapeutic intervention*. Minneapolis: University of Minnesota Press.
- Finn, S. E. (2007). *In our client's shoes: Theory and techniques of therapeutic assessment*. Mahwah, NJ: Erlbaum.
- Finn, S. E., Fischer, C. T., & Handler, L. (2012). *Collaborative/therapeutic assessment: A case-book and guide*. Hoboken, NJ: Wiley.
- Finn, S. E., Moes, E. J., & Kaplan, E. (2001). The consumer's point of view. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 13–46). Lutz, FL: Psychological Assessment Resources.
- Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment*, 4, 278–287.

- Finn, S. E., & Tonsager, M. E. (1997). Information gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment*, 9, 374–385.
- Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools*, 43(8), 835–853.
- First, M. B., Frances, A., Widiger, T. A., Pincus, H. A., & Davis, W. W. (1992). DSM-IV and behavioral assessment. *Behavioral Assessment*, 14, 297–306.
- First, M. B., & Gibbon, M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In M. J. Hilsenroth & A. L. Segal (Eds.) *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 134–143). Hoboken, NJ: Wiley.
- First, M. B., Gibbon, M., Williams, J. B., & Spitzer, R. L. (1996). *Users manual for the Auto SCID-II (for DSM-IV)*. North Tonawanda, NY: Multi-Health Systems.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The Structured Clinical Interview for DSM-III-R personality disorders (SCID-II): Description. *Journal of Personality Disorders*, 9, 83–91.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for Axis I DSM-IV: Disorders Research version—Patient edition (SCID-I/P, Version 2.0)*. New York, NY: New York State Psychiatric Institute, Biometrics Research Department.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)—Clinician version*. Washington, DC: American Psychiatric Press.
- First, M. B., & Tasman, A. (2004). *DSM-IV-TR mental disorders: Diagnosis, etiology, and treatment*. Hoboken, NJ: Wiley.
- First, M. B., Williams, O. B., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5 Disorders (SCID-5)*. Washington, DC: American Psychiatric Press.
- Fischer, J., & Corcoran, K. (2007). *Measures for clinical practice: A sourcebook*. New York, NY: Free Press.
- Fisher, D. C., Beutler, L. E., & Williams, O. B. (1999). Making assessment relevant to treatment planning: The STS Clinician Rating Form. *Journal of Clinical Psychology*, 55, 825–842.
- Fisher, D. C., Ledbetter, M. F., Cohen, N. J., Marmor, D., & Tulskey, D. S. (2000). WAIS-III and WMS-III profiles of mildly to severely brain-injured patients. *Applied Neuropsychology*, 7, 126–132.
- Flanagan, D. P., Alfonso, V. C., Mascolo, J. T., & Hale, J. B. (2011). The Wechsler Intelligence Scale for Children—Fourth Edition in neuropsychological practice. In A. S. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 397–414). New York, NY: Springer.
- Flanagan, D. P., & Harrison, P. L. (Eds.). (2005). *Contemporary intellectual assessment* (2nd ed.). New York, NY: Guilford Press.
- Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.
- Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York, NY: Wiley.
- Flaughner, R. L., & Schrader, W. B. (1978). *Eliminating differentially difficult items as an approach to test bias (RB-78-4)*. Princeton, NJ: Educational Testing Service.
- Flynn, P. M., McCann, J. T., & Fairbank, J. A. (1995). Issues in the assessment of personality disorder and substance abuse using the Millon Clinical Multiaxial Inventory (MCMI-II). *Journal of Clinical Psychology*, 51, 415–421.
- Follette, W. C., & Hayes, S. C. (1992). Behavioral assessment in the DSM era. *Behavioral Assessment*, 14, 293–295.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State. *Journal of Psychiatric Research*, 12, 189–198.
- Folstein, M. F., Romanoski, A. J., Nestadt, G., Chahal, R., Merchant, A., Shapiro, S., ... & McHugh, P. R. (1985). Brief report on the clinical reappraisal of the Diagnostic Interview Schedule carried out at the Johns Hopkins site of the Epidemiological Catchment Area Program of the NIMH. *Psychological Medicine*, 15, 809–814.

- Fonagy, P., Cottrell, D., Phillips, J., Bevington, D., Glaser, D., & Allison, E. (2014). *What works for whom? A critical review of treatments for children and adolescents* (2nd ed.). New York, NY: Guilford Press.
- Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A review and illustration with the MMPI-2 computerized adaptive version. *Psychological Assessment, 19*, 14–24.
- Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*, 248–260.
- Fowler, J. C. (2012). Suicide risk assessment in clinical practice: Pragmatic guidelines for imperfect assessments. *Psychotherapy, 49*(1), 81–90.
- Frank, G. (1990). Research on the clinical usefulness of the Rorschach: The diagnosis of schizophrenia. *Perceptual and Motor Skills, 71*, 573–578.
- Frank, J. D. (1973). *Persuasion and healing: A comparative study of psychotherapy* (rev. ed.). Baltimore, MD: Johns Hopkins University Press.
- Frank, J. D., & Frank, J. B. (2004). Therapeutic components shared by all psychotherapies. In A. Freeman, M. J. Mahoney, P. Devito, & D. Martin (Eds.), *Cognition and psychotherapy* (2nd ed., pp. 45–78). New York, NY: Springer.
- Franzen, M. D. (2004). Behavioral neuropsychology. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3, pp. 386–401). Hoboken, NJ: Wiley.
- Fray, P. J., Robbins, T. W., & Sahakian, B. J. (1996). Neuropsychological applications of CANTAB. *International Journal of Geriatric Psychiatry, 11*, 329–336.
- Free, M. L., & Oei, T. P. S. (1989). Biological and psychological processes in the treatment and maintenance of depression. *Clinical Psychology Review, 9*, 653–688.
- Freilich, B. M., & Hyer, L. A. (2007). Relation of the Repeatable Battery for the assessment of neuropsychological status to measures of daily functioning in dementia. *Psychological Reports, 101*, 119–129.
- Friedman, A. F., Bolinsky, P. K., Levak, R. W., & Nichols, D. S. (2014). *Psychological assessment with the MMPI-2/MMPI-2-RF* (3rd ed.). New York, NY: Routledge/Taylor & Francis.
- Friedman, A. F., Lewak, R., Nichols, D., & Webb, J. F. (2000). *Psychological assessment with the MMPI-2*. Mahwah, NJ: Erlbaum.
- Friman, P. C., Handwerk, M. L., Smith, G. L., Larzelere, R. E., Lucas, C. P., & Shaffer, D. M. (2000). External validity of conduct and oppositional defiant disorders determined by the NIMH Diagnostic Interview Schedule for Children. *Journal of Abnormal Child Psychology, 28*, 277–286.
- Frueh, B. C., & Kinder, B. N. (1994). The susceptibility of the Rorschach Inkblot Test to malin-gering of combat-related PTSD. *Journal of Personality Assessment, 62*, 280–298.
- Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56*, 243–262.
- Fuld, P. A. (1984). Test profile of cholinergic dysfunction and of Alzheimer's-type dementia. *Journal of Clinical Neuropsychology, 6*, 380–392.
- Fuller, A. K., & Blashfield, R. K. (1989). Masochistic personality disorder: A prototype analysis of diagnosis and sex bias. *Journal of Nervous and Mental Disease, 177*(3), 168–172.
- Fyer, A. J., Endicott, J., Manuzza, S., & Klein, D. F. (1985). *Schedule of Affective Disorders and Schizophrenia—Lifetime version (modified for the study of anxiety disorders)*. New York: New York State Psychiatric Institute, Anxiety Disorder Clinic.
- Fyer, A. J., Endicott, J., Manuzza, S., & Klein, D. F. (1995). *Schedule for Affective Disorders and Schizophrenia—Lifetime version (modified for the study of anxiety disorders, updated for DSM-IV; SADS-IV)*. Unpublished measure, New York State Psychiatric Institute, Anxiety Genetics Unit.
- Gale, S. D., Johnson, S. C., Bigler, E. D., & Blatter, D. D. (1995). Nonspecific white matter degeneration following traumatic brain injury. *Journal of the International Neuropsychological Society, 1*, 17–28.

- Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62*, 543–549.
- Gallucci, N. T. (1994). Criteria associated with clinical scales and Harris-Lingoes subscales of the Minnesota Multiphasic Personality Inventory with adolescent inpatients. *Psychological Assessment, 6*, 179–187.
- Gambrill, E. D., & Richey, C. A. (1975). An assertion inventory for use in assessment and research. *Behavior Therapy, 6*, 550–561.
- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4*, 641–655.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.
- Garb, H. N. (1992). The trained psychologist as expert witness. *Clinical Psychology Review, 12*, 451–467.
- Garb, H. N. (1994a). Judgment research: Implications for clinical practice and testimony in court. *Applied and Preventive Psychology, 3*, 173–183.
- Garb, H. N. (1994b). Toward a second generation of statistical prediction rules in psychodiagnostics and personality assessment. *Computers in Human Behavior, 10*, 377–394.
- Garb, H. N. (1998). *Studying the clinician*. Washington, DC: American Psychological Association.
- Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. *Assessment, 6*, 313–317.
- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that. *Psychological Assessment, 12*, 31–39.
- Garb, H. N. (2003). Incremental validity and the assessment of psychopathology in adults: Incremental validity and utility in clinical assessment. *Psychological Assessment, 15*, 508–520.
- Garb, H. N. (2005a). Roots of the Rorschach controversy. *Clinical Psychology Review, 25*, 97–118.
- Garb, H. N. (2005b). Clinical judgment and decision making. *Annual Review of Psychology, 1*, 67–89.
- Garb, H. N. (2007). Computer-administered interviews and rating scales. *Psychological Assessment, 19*, 4–13.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from a meta-analysis. *Psychological Science, 9*, 402–404.
- Garb, H. N., Wood, J. M., Lilienfeld, S. O., & Nezworski, M. T. (2002). Effective use of projective techniques in clinical practice: Let the data help with selection and interpretation. *Professional Psychology: Research and Practice, 33*, 454–463.
- Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment, 13*, 433–448.
- Gardner, H. (2006). *Multiple intelligences: The theory in practice* (2nd ed.). New York, NY: Basic Books.
- Garfield, S. L. (1994). Research on client variables in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 190–228). New York, NY: Wiley.
- Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine, 9*, 273–279.
- Garratt, G., Ingram, R., Rand, K. L., & Sawalani, G. (2007). Cognitive processes in cognitive therapy: Evaluation of the mechanisms of change in the treatment of depression. *Clinical Psychology: Science and Practice, 14*, 224–239.
- Garvey, M. J., Hollon, S. D., & DeRubeis, R. J. (1994). Do depressed patients with higher pretreatment stress levels respond better to cognitive therapy than imipramine? *Journal of Affective Disorders, 32*, 45–50.

- Gass, C. S., & Apple, C. (1997). Cognitive complaints in closed-head injury: Relationship to memory test performance and emotional disturbance. *Journal of Clinical and Experimental Neuropsychology*, 19, 290–299.
- Gass, C. S., & Brown, M. C. (1992). Neuropsychological test feedback to patients with brain dysfunction. *Psychological Assessment*, 4, 272–277.
- Gaw, K. F., & Beutler, L. E. (1995). Integrating treatment recommendations. In L. E. Beutler & M. R. Berren (Eds.), *Integrative assessment of adult personality* (pp. 280–319). New York, NY: Guilford Press.
- Geffken, G. R., Keeley, M. L., Kellison, I., Storch, E. A., & Rodrigue, J. R. (2006). Parental adherence to child psychologists' recommendations from psychological testing. *Professional Psychology: Research & Practice*, 37, 499–505.
- Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology* (Vol. 10, pp. 95–117). Hoboken, NJ: Wiley.
- Geller, J., Cockell, S. J., & Drab, D. L. (2001). Assessing readiness for change in the eating disorder: The psychometric properties of the Readiness and Motivation Interview. *Psychological Assessment*, 13, 189–198.
- George, L., Durbin, J., Sheldon, T., & Goering, P. (2002). Patient and contextual factors related to the decision to hospitalize patients from emergency psychiatric services. *Psychiatric Services*, 53(12), 1586–1591.
- Gerstle, R. M., Geary, D. C., Himelstein, P., & Reller-Geary, L. (1988). Rorschach predictors of therapeutic outcome for inpatient treatment of children: A proactive study. *Journal of Clinical Psychology*, 44, 277–280.
- Gervais, R. O., Ben-Porath, Y. W., Wygant, D. B., & Greene, P. (2007). Development and validation of a response bias scale (RBS) for the MMPI-2. *Assessment*, 14(2), 196–208.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2–RF overreporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology*, 25(4), 274–284.
- Ghassemzadeh, H., Mojtabai, R., Karamghadiri, N., & Ebrahimkhani, N. (2005). Psychometric properties of a Persian-language version of the Beck Depression Inventory–Second Edition: BDI–II–PERSIAN. *Depression and Anxiety*, 21(4), 185–192.
- Giannakou, M., Roussi, P., Kosmides, M., Kiosseoglou, G., Adamopoulou, A., & Garyfallos, G. (2013). Adaptation of the Beck Depression Inventory–II to the Greek population. *Hellenic Journal of Psychology*, 10, 120–146.
- Gibertini, M., Brandenberg, N. A., & Retzlaff, P. D. (1986). The operating characteristics of the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment*, 50, 554–567.
- Gill, H. S. (1966). Delay of response and reaction to color on the Rorschach. *Journal of Projective Techniques and Personality Assessment*, 30, 545–552.
- Gilley, D. W., Wilson, R. S., Fleischmann, D. A., Harrison, D. W., Goetz, C. G., & Tanner, C. M. (1995). Impact of Alzheimer's-type dementia and information source on the assessment of depression. *Psychological Assessment*, 7, 42–48.
- Gilmore, D. C., Beehr, T. A., & Love, K. G. (1986). Effects of applicant sex, applicant physical attractiveness, and type of job on interview decisions. *Journal of Occupational Psychology*, 59, 103–109.
- Ginting, H., Näring, G., van der Veld, W. M., Srisayekti, W., & Becker, E. S. (2013). Validating the Beck Depression Inventory–II in Indonesia's general population and coronary heart disease patients. *International Journal of Clinical and Health Psychology*, 13(3), 235–242.
- Glass, C. R., & Merluzzi, T. V. (2000). Cognitive and behavioral assessment. In C. E. Watkins & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (pp. 175–224). Mahwah, NJ: Erlbaum.
- Glass, C. R., Merluzzi, T. V., Biever, J. L., & Larsen, K. H. (1982). Cognitive assessment of social anxiety: Development and validation of a self-statement questionnaire. *Cognitive Therapy and Research*, 6, 37–55.

- Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement. *Journal of Special Education, 40*, 103–114.
- Gocka, E. (1965). *American Lake norms for 200 MMPI scales*. Unpublished raw data.
- Godber, T., Anderson, V., & Bell, R. (2000). The measurement and diagnostic utility of intra-subtest scatter in pediatric neuropsychology. *Journal of Clinical Psychology, 56*, 101–112.
- Gogos, A., Joshua, N., & Rossell, S. L. (2010). Use of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) to investigate group and gender differences in schizophrenia and bipolar disorder. *Australian and New Zealand Journal of Psychiatry, 44*(3), 220–229.
- Gold, J. M., Goldberg, R., MnNary, S., Dixon, L., & Lehman, A. (2002). Cognitive correlates of job tenure among patients with severe mental illness. *American Journal of Psychiatry, 159*, 1395–1402.
- Goldbeck, L., Daseking, M., Hellwig-Brida, S., Waldmann, H. C., & Petermann, F. (2010). Sex differences on the German Wechsler Intelligence Test for Children (WISC-IV). *Journal of Individual Differences, 31*(1), 22–28.
- Goldberg, E., & Bilder, R. M. (1987). The frontal lobes and heirarchical organization of cognitive control. In E. Perecman (Ed.), *The frontal lobes revisited* (pp. 159–187). New York, NY: IRBN Press.
- Golden, C. J. (1976). The identification of brain damage by an abbreviated form of the Halstead-Reitan Neuropsychological Battery. *Journal of Clinical Psychology, 32*, 821–826.
- Golden, C. J., Purisch, A. D., & Hammeke, T. A. (1985). *Luria-Nebraska Neuropsychological Battery: Forms I and II (manual)*. Los Angeles, CA: Western Psychological Services.
- Goldfried, M. R. (1982). On the history of therapeutic integration. *Behavior Therapy, 13*, 572–593.
- Goldfried, M. R. (1983). A behavior therapist looks at reapproachment. *Journal of Humanistic Psychology, 23*, 97–107.
- Goldstein, A. M. (2007). *Forensic psychology: Emerging topics and expanding roles*. Hoboken, NJ: Wiley.
- Goncalves, A. A., Woodward, M. J., & Millon, T. (1994). The Millon Clinical Multiaxial Inventory–II. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 161–184). Hillsdale, NJ: Erlbaum.
- Gooding, P. (2013). Supported decision-making: A rights-based disability concept and its implications for mental health law. *Psychiatry, Psychology and Law, 20*(3), 431–451.
- Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination (BDAE)*. Philadelphia, PA: Lea & Febiger.
- Gooding, P. (2013). Supported decision-making: A rights-based disability concept and its implications for mental health law. *Psychiatry, Psychology and Law, 20*(3), 431–451.
- Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Posttraumatic stress disorder in veterans: The utility of the MMPI-2–RF validity scales in detecting overreported symptoms. *Psychological Assessment, 25*(3), 671–678.
- Gordon, N. G., & Swart, E. C. (1973). A comparison of the Harris-Lingoes subscales between the original standardization population and an inpatient Veterans Administration hospital population. *Newsletter for Research in Mental Health and Behavioral Sciences, 15*, 28–31.
- Gorwood, P., Corruble, E., Falissard, B., & Goodwin, G. M. (2008). Toxic effects of depression on brain function: Impairment of brain recall and the cumulative length of depressive disorder in a sample of depressed outpatients. *American Journal of Psychiatry, 165*, 731–739.
- Gotts, E. E., & Knudson, T. E. (2005). *The clinical interpretation of the MMPI-2: A content cluster approach*. Mahwah, NJ: Erlbaum.
- Gough, H. G., & Heilbrun, A. B., Jr., (1983). *Adjective Check List manual*. Palo Alto, CA: Consulting Psychologists Press.
- Gould, R. A., Otto, M. W., & Pollack, M. H. (1995). A meta-analysis of treatment outcome for panic disorder. *Clinical Psychology Review, 15*, 819–844.

- Grace, J., Stout, J., & Malloy, P. F. (1999). Assessing frontal lobe behavioral syndromes with the Frontal Lobe Personality Scale. *Assessment*, 6, 269–284.
- Graham, J. R. (1987). *The MMPI: A practical guide* (2nd ed.). New York, NY: Oxford University Press.
- Graham, J. R. (2011). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.
- Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). *MMPI-2 correlates for outpatient mental health settings*. Minneapolis: University of Minnesota Press.
- Graham, J. R., & Lilly, R. S. (1984). *Psychological testing*. Englewood Cliffs, NJ: Prentice-Hall.
- Graham, J. R., Smith, R. L., & Schwartz, G. F. (1986). Stability of MMPI configurations for psychiatric inpatients. *Journal of Consulting and Clinical Psychology*, 49, 477–484.
- Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, 57(2), 264–277.
- Graham, P., & Rutter, M. (1968). The reliability and validity of the psychiatric assessment of the child. II: Interview with the parent. *British Journal of Psychiatry*, 114, 581–592.
- Graves, R. E., Carswell, L. M., & Snow, W. G. (1999). An evaluation of the sensitivity of pre-morbid IQ estimators for detecting cognitive decline. *Psychological Assessment*, 11, 29–38.
- Green, A., Garrick, T., Sheedy, D., Blake, H., Shores, E. A., & Harper, C. (2010). The effect of moderate to heavy alcohol consumption on neuropsychological performance as measured by the Repeatable Battery for the Assessment of Neuropsychological Status. *Alcoholism: Clinical and Experimental Research*, 34(3), 443–450.
- Green, K., Worden, B., Menges, D., McCrady, B., Hunsley, J., & Mash, E. J. (2008). Alcohol use disorders. In J. Hunsley & E. J. Mash (Eds.), *A Guide to Assessments that Work* (pp. 339–369). New York, NY: Oxford University Press.
- Green, R. E., Melo, B., Christensen, B., Ngo, L. A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading. *Journal of Clinical and Experimental Neuropsychology*, 30, 163–172.
- Green, S. B., & Kelley, C. K. (1988). Racial bias in prediction with the MMPI for a juvenile delinquent population. *Journal of Personality Assessment*, 52, 263–275.
- Greenbaum, P. F., Prange, M. E., Friedman, R. M., & Silver, S. E. (1991). Substance abuse prevalence and comorbidity with other psychiatric disorders among adolescents with severe emotional disorders. *Journal of the Academy of Child and Adolescent Psychology*, 30, 575–583.
- Greenberg, L. (2008). Emotion and cognition in psychotherapy: The transforming power of affect. *Canadian Psychology/Psychologie Canadienne*, 49(1), 49–59.
- Greene, R. L. (1987). Ethnicity and MMPI performance: A review. *Journal of Consulting and Clinical Psychology*, 55, 497–512.
- Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual*. Boston, MA: Allyn & Bacon.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston, MA: Allyn & Bacon.
- Greene, R. L. (2010). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Minneapolis, MN: Pearson.
- Greene, R. L. (2011). Some considerations for enhancing psychological assessment. *Journal of Personality Assessment*, 93(3), 198–203.
- Greene, R. L., & Clopton, J. R. (1994). Minnesota Multiphasic Personality Inventory–2. In M. E. Maruish (Ed.), *The use of psychological testing for treatment, planning and outcome assessment* (pp. 137–159). Hillsdale, NJ: Erlbaum.
- Greene, R. L., Weed, N. C., Butcher, J. N., Arredondo, R., & Davis, H. G. (1992). A cross validation of MMPI-2 substance abuse scales. *Journal of Personality Assessment*, 58, 405–410.
- Greenspan, S., & Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 131–150). New York, NY: Guilford Press.

- Greenway, P., & Milne, L. (1999). Relationship between psychopathology, learning disabilities, or both and WISC-III subtest scatter in adolescents. *Psychology in the Schools*, 36, 103–108.
- Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in the detection of noncredible brain injury claims. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 210–235). New York, NY: Guilford Press.
- Gresham, F. M. (1984). Behavioral interviews in school psychology: Issues in psychometric adequacy and research. *School Psychology Review*, 13, 17–25.
- Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO-PI-R. *Personality and Individual Differences*, 36(7), 1545–1558.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75–111.
- Grillo, J., Brown, R. S., Hilsabeck, R., Price, J. R., & Lees-Haley, P. (1994). Raising doubts about claims of malingering: Implications of relationships between MCMI-II and MMPI-2 performances. *Journal of Clinical Psychology*, 50, 651–655.
- Gronnerod, C. (2004). Rorschach assessment of changes following psychotherapy: A meta-analytic review. *Journal of Personality Assessment*, 83, 256–276.
- Gronnerod, C. (2006). Reanalysis of the Gronnerod (2003) Rorschach temporal stability meta-analysis data set. *Journal of Personality Assessment*, 86, 222–225.
- Grossman, L. S., Wasyliw, O. E., Benn, A. F., & Gyoerkoe, K. L. (2002). Can sex offenders who minimize on the MMPI conceal psychopathology on the Rorschach? *Journal of Personality Assessment*, 78, 484–501.
- Grossman, S., & del Rio, C. (2005). The MCMI-III facet scales. In Craig, R. J. (Ed.) *New directions in interpreting the Millon Clinical Multiaxial Inventory-III (MCMI-III)*. Hoboken, NJ: Wiley.
- Grothe, K. B., Dutton, G. R., Jones, G. N., Bodenlos, J., Ancona, M., & Brantley, P. J. (2005). Validation of the Beck Depression Inventory-II in a low-income African American sample of medical outpatients. *Psychological Assessment*, 17, 110–114.
- Groth-Marnat, G. (1985). Evaluating and using psychological testing software. *Human Resource Management Australia*, 23, 16–21.
- Groth-Marnat, G. (1988). A survey of the current and future direction of professional psychology in acute general hospitals in Australia. *Australian Psychologist*, 23, 39–43.
- Groth-Marnat, G. (1992). Past cultural traditions of therapeutic metaphor. *Psychology: An International Journal of Human Behavior*, 29, 1–8.
- Groth-Marnat, G. (1993). Neuropsychological effects of styrene exposure: A review of current literature. *Perceptual and Motor Skills*, 77, 1139–1149.
- Groth-Marnat, G. (1999). Financial efficacy of clinical assessment: Rational guidelines and issues for future research. *Journal of Clinical Psychology*, 55, 813–824.
- Groth-Marnat, G. (Ed.). (2000a). *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration*. New York, NY: Wiley.
- Groth-Marnat, G. (2000b). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology*, 56, 349–365.
- Groth-Marnat, G. (2001). Learning disabilities assessment with the Wechsler intelligence scales (WAIS-III/WISC-III). In A. Kaufman & N. Kaufman (eds.), *Learning disabilities assessment: Psychological assessment and evaluation* (pp. 29–52). Cambridge, UK: Cambridge University Press.
- Groth-Marnat, G. (2003). The psychological report. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 812–817). Thousand Oaks, CA: Sage.
- Groth-Marnat, G. (2009). The five assessment issues you meet when you go to heaven. *Journal of Personality Assessment*, 91(4), 303–310.
- Groth-Marnat, G., & Davis, A. (2014). *Psychological report writing assistant*. Hoboken, NJ: Wiley.

- Groth-Marnat, G., & Edkins, G. (1996). Professional psychologists in general health care settings: A review of the financial efficacy of direct treatment interventions. *Professional Psychology: Research and Practice*, 27, 161–174.
- Groth-Marnat, G., Edkins, G., & Schumaker, J. F. (1995). Psychologists in disease prevention and health promotion: A review of the cost-effectiveness literature. *Psychology*, 32, 127–135.
- Groth-Marnat, G., Gallagher, R. E., Hale, J. B., & Kaplan, E. (2000). The Wechsler intelligence scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A practical guide to test interpretation and integration* (pp. 129–194). New York, NY: Wiley.
- Groth-Marnat, G., Gottheil, E., Liu, W., Clinton, D., & Beutler, L. E. (2008). Personality and treatment planning for psychotherapy: The Systematic Treatment Selection Model. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Handbook of personality and testing* (pp. 620–634). Thousand Oaks, CA: Sage.
- Groth-Marnat, G., & Horvath, L. S. (2006). The psychological report: A review of current controversies. *Journal of Clinical Psychology*, 62(1), 73–81.
- Groth-Marnat, G., Roberts, R., & Beutler, L. E. (2001). Client characteristics and psychotherapy: Perspectives, support, interactions, and implications. *Australian Psychologist*, 36, 115–121.
- Groth-Marnat, G., & Schumaker, J. (1989). Computer-based psychological testing: Issues and guidelines. *American Journal of Orthopsychiatry*, 59, 257–263.
- Groth-Marnat, G., & Teal, M. (2000). Block Design as a measure of everyday spatial ability: A study of ecological validity. *Perceptual and Motor Skills*, 90, 522–526.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A metaanalysis. *Psychological Assessment*, 12, 19–30.
- Groves, J. A., & Engel, R. R. (2007). The German adaptation and standardization of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment*, 88(1), 49–56.
- Guevremont, G. C., & Spiegler, M. D. (1990, November). *What do behavior therapists really do? A survey of the clinical practice of AABT members. Paper presented at the 24th annual convention of the Association for Advancement of Behavior Therapy*, San Francisco, CA.
- Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for white and black children from the standardization sample. *Journal of Educational Psychology*, 73, 227–231.
- Gynther, M. D., & Green, S. B. (1980). Accuracy may make a difference, but does a difference make for accuracy? A response to Pritchard and Rosenblatt. *Journal of Consulting and Clinical Psychology*, 48, 268–272.
- Haaga, D. A., Dyck, M. J., & Ernst, D. (1991). Empirical status of cognitive theory of depression. *Psychological Bulletin*, 110, 215–236.
- Haas, A. P., Hendin, H., & Singer, P. (1987). Psychodynamic and structured interviewing: Issues of validity. *Comprehensive Psychiatry*, 28, 40–53.
- Haber, J. C., & Baum, L. J. (2014). Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) scales as predictors of psychiatric diagnoses. *South African Journal of Psychology*, 44(4), 439–453.
- Hain, J. D. (1964). The Bender Gestalt Test: A scoring method for identifying brain damage. *Journal of Consulting Psychology*, 28, 34–40.
- Hale, J. B., Wycoff, K. L., & Fiorello, C. A. (2011). RTI and cognitive hypothesis testing for specific learning disabilities identification and intervention: The best of both worlds. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 173–202). Hoboken, NJ: Wiley.
- Hale, W. D., Cochran, C. D., & Hedgepeth, B. E. (1984). Norms for the elderly on the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology*, 52, 321–322.
- Haley, W. E., Larson, D. G., Kasl-Godley, J., Neimeyer, R. A., & Kwilosz, D. M. (2003). Roles for psychologists in end-of-life care: Emerging models of practice. *Professional Psychology: Research and Practice*, 34(6), 626–633.
- Hall, C. (1983, December 6). Psychiatrist's computer use stirs debate. *The Wall Street Journal*, pp. 35, 39.

- Hall, G. C. N. (2001). Psychotherapy research with ethnic minorities: Empirical, ethical, and conceptual issues. *Journal of Consulting and Clinical Psychology*, 69(3), 502–510.
- Hall, G. C. N., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment*, 11, 186–197.
- Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment*, 75, 280–294.
- Hammen, C. L. (1978). Depression, distortion, and life stress in college students. *Cognitive Therapy and Research*, 2, 189–192.
- Hammen, C. L., & Krantz, S. (1976). Effect of success and failure on depressive cognitions. *Journal of Abnormal Psychology*, 85, 577–586.
- Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (2009). *Comprehensive Test of Nonverbal Intelligence* (2nd ed.). Austin, TX: PRO-ED.
- Han, K., Park, H. I., Weed, N. C., Lim, J., Johnson, A., & Joles, C. (2013). Gender differences on the MMPI across American and Korean adult and adolescent normative samples. *Journal of Personality Assessment*, 95(2), 197–206.
- Handel, R. W., & Archer, R. P. (2008). An investigation of the psychometric properties of the MMPI-2 restructured clinical (RC) scales with mental health patients. *Journal of Personality Assessment*, 90, 239–249.
- Handel, R. W., & Ben-Porath, Y. S. (2000). Multicultural assessment with the MMPI-2: Issues for research and practice. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 229–245). Mahwah, NJ: Erlbaum.
- Handelsman, M. M., & Galvin, M. D. (1988). Facilitating informed consent for outpatient psychotherapy: A suggested written format. *Professional Psychology: Research and Practice*, 19, 223–225.
- Hansell, A. G., Lerner, H. D., Milden, R. S., & Ludolph, P. (1988). Single-sign Rorschach suicide indicators: A validity study using a depressed inpatient population. *Journal of Personality Assessment*, 52, 658–669.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessment for sex offenders* (User Rep No.1999–02). Ottawa, Ontario: Department of the Solicitor General of Canada.
- Hara, N., Nishimura, Y., Yokoyama, C., Inoue, K., Nishida, A., Tanii, H., ... Okazaki, Y. (2012). The development of agoraphobia is associated with the symptoms and location of the patient's first panic attack. *BioPsychoSocial Medicine*, 6(1), 1–8.
- Harding, C. M., & Strauss, J. S. (2007). The course of schizophrenia: An evolving concept. In M. Alpert (Ed.), *Controversies in Schizophrenia: Changes and Constancies: Proceedings of the 74th Annual Meeting of the American Psychological Association* (pp. 339–350). New York, NY: Guilford Press.
- Hargrave, G. E., Hiatt, D., Ogard, E. M., & Karr, C. (1994). Comparison of the MMPI and MMPI-2 for a sample of peace officers. *Psychological Assessment*, 6, 27–32.
- Harkness, A. R., McNulty, J. L., Ben-Porath, Y. S., & Graham, J. R. (2002). *MMPI-2 Personality Psychopathology Five (PSY-5) Scales: Gaining an overview for case conceptualization and treatment planning*. Minneapolis: University of Minnesota Press.
- Harrell, T. H., Honaker, L. M., & Parnell, T. (1992). Equivalence of the MMPI-2 with the MMPI in psychiatric patients. *Psychological Assessment*, 4, 460–465.
- Harris, R. E., & Lingo, J. C. (1968). *Subscales for the MMPI: An aid to profile interpretation*. Los Angeles, CA: University of California, Department of Psychiatry. (Original work published 1955)
- Hartshorne, H., & May, M. A. (1928). *Studies in deceit*. New York, NY: Macmillan.
- Harvey, V. S. (1997). Improving readability of psychological reports. *Professional Psychology: Research and Practice*, 28, 271–274.
- Harvey, V. S. (2006). Variables affecting the clarity of reports. *Journal of Clinical Psychology*, 62, 5–18.

- Harwood, T. M., Beutler, L. E., & Groth-Marnat, G. (2011). *Integrative assessment of adult personality* (3rd ed.). New York, NY: Guilford Press.
- Harwood, T. M., & Williams, O. (2003). Identifying treatment relevant assessment: The STS. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (pp. 65–81). New York, NY: Guilford Press.
- Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*, 35(5), 485–491.
- Hathaway, S. R., & McKinley, J. C. (1983). *Manual for the administration and scoring of the MMPI*. Minneapolis, MN: National Computer Systems.
- Hathaway, S. R., & Monachesi, E. D. (1963). *Adolescent personality and behavior: MMPI patterns of normal, delinquent, dropout, and other outcomes*. Minneapolis: University of Minnesota Press.
- Hauff, E., & Vaglum, P. (1994). Chronic posttraumatic stress disorder in Vietnamese refugees. A prospective community study of prevalence, course, psychopathology, and stressors. *Journal of Nervous and Mental Disease*, 182, 85–90.
- Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses. *Journal of Counseling Psychology*, 40, 303–315.
- Hawkins, K. A. (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III Technical Manual clinical sample data: A preliminary approach to clinical utility. *Clinical Neuropsychologist*, 12, 535–551.
- Hawkins, K. A., Sullivan, T. E., & Choi, E. J. (1997). Memory defects in schizophrenia: Inadequate assimilation or true amnesia? Findings from the Wechsler Memory Scale–Revised. *Journal of Psychiatry and Neuroscience*, 22, 169–179.
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107, 311–327.
- Hayes, J. A. (1997). What does the Brief Symptom Inventory measure in college and university counseling center clients? *Journal of Counseling Psychology*, 44, 360–367.
- Haynes, S. N. (1995). Introduction to the special section in chaos theory and psychological assessment. *Psychological Assessment*, 7, 3–4.
- Haynes, S. N. (2006). Psychometric considerations. In M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 17–42). New York, NY: Elsevier.
- Haynes, S. N., & Heiby, E. M. (Eds.). (2004). *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3). Hoboken, NJ: Wiley.
- Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment*. New York, NY: Kluwer Academic.
- Haynes, S. N., O'Brien, W., & Kaholokula, J. (2011). *Behavioral assessment and case formulation*. Hoboken, NJ: Wiley.
- Haynes, S. N., & Yoshioka, D. T. (2007). Clinical assessment applications of ambulatory sensors. *Psychological Assessment*, 19, 44–57.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted for neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources.
- Heaton, R. K., Taylor, M. J., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In D. S. Tulsky, D. H. Saklofske, G. J. Chelune, R. K. Heaton, & R. J. Ivnik, *Clinical interpretation of the WAIS-III and WMS-III* (pp. 183–204). San Diego, CA: Academic Press.
- Hebben, N., & Milberg, W. (2002). *Essentials of neuropsychological assessment*. New York, NY: Wiley.
- Hedlund, J. L., Sletten, I. W., Evenson, R. C., Altman, H., & Cho, D. W. (1977). Automated psychiatric information systems: A critical review of Missouri's Standard System of Psychiatry (SSOP). *Journal of Operational Psychiatry*, 8, 5–26.
- Heilbrun, K., Marczyk, G. G., & Dematteo, D. (2002). *Forensic mental health assessment: A casebook*. New York, NY: Oxford University Press.

- Heimberg, R. G. (1994). Cognitive assessment strategies and the measurement of outcome of treatment for social phobia. *Behavior Research and Therapy*, 32, 269–280.
- Heimberg, R. G., Harrison, D. F., Goldberg, L. S., Desmarais, S., & Blue, S. (1979). The relationship of self-report and behavioral assertion in an offender population. *Journal of Behavior Therapy and Experimental Psychiatry*, 10, 283–286.
- Hellkamp, D. T., & Hogan, M. E. (1985). Differentiation of organics from functional psychiatric patients across various I.Q. ranges using the Bender-Gestalt and Hutt scoring system. *Journal of Clinical Psychology*, 41, 259–264.
- Helmes, E. (2000). Learning and memory. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 293–334). New York, NY: Wiley.
- Helmes, E., & Reddon, J. R. (1993). A perspective on development in assessing psychopathology: A critical review of MMPI and MMPI-2. *Psychological Bulletin*, 113, 453–471.
- Helms, J. L. (2003). Analysis of Miranda reading levels across jurisdictions: Implications for evaluating waiver competency. *Journal of Forensic Psychology Practice*, 3, 25–37.
- Helzer, J. E., & Robins, L. N. (1988). The Diagnostic Interview Schedule: Its development, evolution and use. *Social Psychiatry and Psychiatric Epidemiology*, 23, 6–16.
- Helzer, J. E., Robins, L. N., Croughan, J. L., & Welner, A. (1981). Renard Diagnostic Interview: Its reliability and procedural validity with physicians and lay interviewers. *Archives of General Psychiatry*, 38, 393–398.
- Helzer, J. E., Robins, L. N., McEvoy, L. F., Spitznagel, E. L., Stolzman, R. K., Farmer, A., & Brockington, I. F. (1985). A comparison of Clinical and Diagnostic Interview Schedule diagnoses: Physician re-examination of lay-interviewed cases in the general population. *Archives of General Psychiatry*, 42, 657–666.
- Helzer, J. E., Spitznagel, E. L., & McEvoy, L. (1987). The predictive validity of lay Diagnostic Interview Schedule diagnoses in the general population: A comparison with physician examiners. *Archives of General Psychiatry*, 44, 1069–1077.
- Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the “remembrance of things past”: A longitudinal evaluation of the retrospective method. *Psychological Assessment*, 6, 92–101.
- Herjanic, B., & Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology*, 51, 127–134.
- Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement on diagnosis comparing child and patient interviews. *Journal of Abnormal Psychology*, 10, 325–336.
- Hersen, M. (Ed.). (2005a). *Clinician's handbook of adult behavioral assessment*. New York, NY: Elsevier.
- Hersen, M. (Ed.). (2005b). *Clinician's handbook of child behavioral assessment*. New York, NY: Elsevier.
- Hersen, M., & Bellack, A. S. (Eds.). (2002). *Dictionary of behavioral assessment techniques* (2nd ed). New York, NY: Pergamon Press.
- Hersen, M., & Greaves, S. T. (1971). Rorschach productivity as related to verbal performance. *Journal of Personality Assessment*, 35, 436–441.
- Hersen, M., & Sturmey, P. (2012). *Handbook of evidence-based practice in clinical psychology* (Vols. 1 and 2). Hoboken, NJ: Wiley.
- Hertz, M. R. (1948). Suicidal configurations in Rorschach records. *Rorschach Research Exchange*, 12, 3–58.
- Hertz, M. R. (1976). Detection of suicidal risks with the Rorschach. In M. Abt & S. L. Weissman (Eds.), *Acting out: Theoretical and clinical aspects* (2nd ed.). New York, NY: Aronson.
- Herzog, D. B., Keller, M. B., Sacks, N. R., Yeh, C. J., & Lavori, P. W. (1992). Psychiatric comorbidity in treatment-seeking anorexics and bulimics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 810–818.
- Hesse, M., Guldager, S., & Holm Linneberg, I. (2012). Convergent validity of MCMI-III clinical syndrome scales. *British Journal of Clinical Psychology*, 51(2), 172–184.

- Higgins, R. L., Alonso, R. R., & Pendleton, M. G. (1979). The validity of roleplay assessments of assertiveness. *Behavior Therapy, 10*, 655–662.
- Hill, J. S., Pace, T. M., & Robbins, R. R. (2010). Decolonizing personality assessment and honoring indigenous voices: A critical examination of the MMPI-2. *Cultural Diversity and Ethnic Minority Psychology, 16*(1), 16–25.
- Hiller, J. B., Rosenthal, R., Bornstein, R. A., Berry, D. T. R., & Brunnell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*, 27–66.
- Hilsenroth, M. J., & Handler, L. (1995). A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. *Journal of Personality Assessment, 64*, 243–257.
- Hilsenroth, M. J., & Stricker, G. (2004). A consideration of challenges to psychological assessment instruments used in forensic settings. *Journal of Personality Assessment, 83*, 141–152.
- Hiroe, T., Kojima, M., Yamamoto, I., Nojima, S., Kinoshita, Y., Hashimoto, N., ... Furukawa, T. A. (2005). Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory–II in Japanese patients with depression. *Psychiatry Research, 135*(3), 229–235.
- Hishinuma, E. S., Miyamoto, R. H., Nishimura, S. T., & Nahulu, L. B. (2000). Differences in State–Trait Anxiety Inventory scores for ethnically diverse adolescents in Hawaii. *Cultural Diversity and Ethnic Minority Psychology, 6*(1), 73–83.
- Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*(1), 52–88.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469–477.
- Hoge, R. D., Andrews, D. A., Robinson, D., & Hollett, J. (1988). The construct validity of interview-based assessments in family counseling. *Journal of Clinical Psychology, 44*, 563–571.
- Holdnack, J. A., Drozdick, L., Iverson, G. L., & Weiss, L. G. (2013). *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation*. San Diego, CA: Elsevier Science.
- Holdnack, J. A., Schoenberg, M. R., Lange, R. T., & Iverson, G. L. (2013). Predicting premorbid ability for WAIS–IV, WMS–IV and WASI–II. In J. A. Holdnack, L. Drozdick, G. L. Iverson, & L. G. Weiss (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation*. San Diego, CA: Elsevier Science.
- Holland, A. L. (1980). *Communicative abilities in daily living: A test of functional communication for aphasic adults*. Austin, TX: PRO-ED.
- Hollis, J. W., & Donna, P. A. (1979). *Psychological report writing: Theory and practice*. Muncie, IN: Accelerated Development.
- Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research, 4*, 383–395.
- Holmes, C. B., Dungan, D. S., & Medlin, W. J. (1984). Reassessment of inferring personality traits from Bender-Gestalt drawing styles. *Journal of Clinical Psychology, 40*, 1241–1243.
- Holmes, C. B., & Stephens, C. L. (1984). Consistency of edging on the Bender-Gestalt, Memory for Designs, and Draw-A-Person Tests. *Journal of Psychology, 117*, 269–271.
- Holtzman, W. H. (1988). Beyond the Rorschach. *Journal of Personality Assessment, 52*, 578–609.
- Hoofien, D., Barak, O., Vakil, E., & Gilboa, A. (2005). Symptom Checklist–90 Revised scores in persons with traumatic brain injury: Affective reactions or neurobehavioral outcomes of the injury? *Applied Neuropsychology, 12*, 30–39.
- Hope, D. A., & Heimberg, R. G. (1993). Social phobia and social anxiety. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 99–136). New York, NY: Guilford Press.
- Hopwood, C. J., Flato, C. G., Ambwani, S., Garland, B. H., & Morey, L. C. (2009). A comparison of Latino and Anglo socially desirable responding. *Journal of Clinical Psychology, 65*(7), 769–780.

- Hopwood, C. J., & Moser, J. S. (2011). Personality Assessment Inventory internalizing and externalizing structure in college students: Invariance across sex and ethnicity. *Personality and Individual Differences*, 50(1), 116–119.
- Hopwood, C. J., & Richard, D. C. S. (2005). Graduate student WAIS-III scoring accuracy is a function of full scale IQ and complexity of examiner tasks. *Assessment*, 12(4), 445–454.
- Horowitz, M. J. (1985). *Report of the program on conscious and unconscious mental processes of the John D. and Catherine T. MacArthur Foundation*. San Francisco: University of California.
- Horvath, A. O., & Goheen, M. D. (1990). Factors mediating the success of defiance- and compliance-based interventions. *Journal of Consulting and Clinical Psychology*, 37, 363–371.
- Houts, A. C. (2002). Discovery, invention, and the expansion of the modern *Diagnostic and Statistical Manuals of Mental Disorders*. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 17–69). Washington, DC: American Psychological Association.
- Hoyt, M. F. (1994). Single session solutions. In M. Hoyt (Ed.), *Constructive therapies* (pp. 140–159). New York, NY: Guilford Press.
- Hser, Y., Evans, E., Grella, C., Ling, W., & Anglin, D. (2015). Long-term course of opioid addiction. *Harvard Review of Psychiatry*, 23(2), 76–89.
- Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 28(2), 192–218.
- Huffcutt, A. I., & Arthur, W., Jr., (1994). Hunter and Hunter (1994) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–194.
- Humphrey, D. H., & Dahlstrom, W. G. (1995). The impact of changing from the MMPI to the MMPI-2 on profile configurations. *Journal of Personality Assessment*, 64, 428–439.
- Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach? *Psychological Assessment*, 13, 472–485.
- Hunsley, J., Hanson, R. K., & Parker, K. C. H. (1988). A summary of the reliability and stability of MMPI scales. *Journal of Clinical Psychology*, 44, 44–46.
- Hunsley, J., & Mash, E. J. (Eds.). (2008). *A guide to assessments that work*. New York, NY: Oxford University Press.
- Hunsley, J., & Meyer, G. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, 15, 446–455.
- Hunter, E. E., Powell, B. J., Penick, E. C., Nickel, E. J., Othmer, E., & DeSouza, C. (2000). Development and validation of a mania scale for the Symptom Checklist-90-R. *Journal of Nervous and Mental Disease*, 188, 176–179.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job employment. *Psychological Bulletin*, 96, 72–98.
- Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, 2, 447–472.
- Hunter, J. E., & Schmidt, F. L. (2000). Racial and gender bias in ability and achievement tests. *Psychology, Public Policy, and Law*, 6, 151–158.
- Hutt, M. L., & Briskin, G. J. (1960). *The clinical use of the revised Bender-Gestalt Test*. New York, NY: Grune & Stratton.
- Hwang, W. C. (2009). The Formative Method for Adapting Psychotherapy (FMAP): A community-based developmental approach to culturally adapting therapy. *Professional Psychology: Research and Practice*, 40(4), 361–368.
- Iacano, W. G., & Patrick, C. J. (2006). Polygraph (“lie detector”) testing: Current status and emerging trends. In I. B. Weiner & A. K. Hess (Eds.), *The handbook of forensic psychology* (3rd ed., pp. 552–588). Hoboken, NJ: Wiley.

- Ingram, R. E., Kendall, P. C., Siegle, G., Guarino, J., & McLaughlin, S. C. (1995). Psychometric properties of the Positive Automatic Thoughts Questionnaire. *Psychological Assessment*, 7, 495–507.
- Iverson, G. L., Franzen, M. D., & Hammond, J. A. (1995). Examination of inmate's ability to malingering on the MMPI-2. *Psychological Assessment*, 7, 115–117.
- Jackson, D. N., Fraboni, M., & Helms, E. (1997). MMPI-2 content scales: How much content do they measure? *Assessment*, 4, 111–117.
- Jackson, K. M., & Trull, T. J. (2001). The factor structure of the Personality Assessment Inventory–Borderline Features (PAI–BOR) scale in a nonclinical sample. *Journal of Personality Disorders*, 15(6), 536–545.
- Jacob, R. G., Turner, S. M., Szekely, B. C., & Eidelman, B. H. (1983). Predicting outcome of relaxation therapy in headaches: The role of “depression.” *Behavior Therapy*, 14, 457–465.
- Jacobson, N. S., Follette, W. C., & Pagel, M. (1986). Predicting who will benefit from behavioral marital therapy. *Journal of Consulting and Clinical Psychology*, 54, 518–522.
- Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, 74(6), 1556–1565.
- Jankowski, D. (2002). *A beginner's guide to the MCMI-III*. Washington, DC: American Psychological Association.
- Janus, M., Tolbert, H., Caestro, K., & Toepfer, S. (1996). Clinical accuracy ratings of MMPI approaches for adolescents: Adding ten years and the MMPI-A. *Journal of Personality Assessment*, 67, 364–383.
- Jarvis, P. E., & Barth, J. (1994). *The Halstead-Reitan Neuropsychological Battery: A guide to interpretation and clinical application*. Odessa, FL: Psychological Assessment Resources.
- Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnosis for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology*, 70, 158–168.
- Jensen, A. R. (1965). Review of the Rorschach. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (pp. 501–509). Highland Park, NJ: Gryphon Press.
- Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, 39, 1–23.
- Jensen, A. R. (1972). *Genetics and education*. New York, NY: Harper & Row.
- Jensen, A. R. (1984). The black-white difference on the K–ABC: Implications for future tests. *Journal of Special Education*, 18, 377–408.
- Jensen, A. R., & Rohwer, W. D. (1966). The Stroop Color-Word Test: A review. *Acta Psychologica*, 25, 36–93.
- Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology*, 76(5), 711–722.
- Johnson, J. H., Null, C., Butcher, J. M., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology*, 47, 105–114.
- Johnson, J. W., & Williams, T. A. (1980). Using on-line computer technology to improve service response and decision-making effectiveness in a mental health admitting system. In J. B. Sidowski & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 237–252). Norwood, NJ: Ablex.
- Johnson, L. C., Murphy, S. A., & Dimond, M. (1996). Reliability, construct validity, and subscale norms of the Brief Symptom Inventory when administered to bereaved parents. *Journal of Nursing Measurement*, 4, 117–127.
- Johnson, M. H., Margo, P. A., & Stern, S. L. (1986). Use of the SADS–C as a diagnostic and symptom severity measure. *Journal of Consulting and Clinical Psychology*, 54, 546–551.
- Jones, M. C. (1924). The elimination of children's fears. *Journal of Experimental Psychology*, 7, 382–390.
- Jones, R. G. (1969). A factored measure of Ellis's Irrational Belief System. *Dissertation Abstracts International*, 29(43), 4379B–4380B (UMI No. 69–64).

- Jongsma, A. E., Jr., Peterson, L. M., & Bruce, T. J. (2006). *The complete adult psychotherapy treatment planner* (4rd ed.). Hoboken, NJ: Wiley.
- Jongsma, A. E., Peterson, L. M., & Bruce, T. J. (2014). *The complete adult psychotherapy treatment planner: Includes DSM-5 updates* (5th ed.). Hoboken, NJ: Wiley.
- Jorgensen, K., Anderson, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment*, 7, 259–280.
- Jung, S. (2014). *Ecological validity of traditional neuropsychological tests: Role of memory and executive skills in predicting functional ability in a clinical population*. Retrieved from <http://ddp.policyresearchinc.org/wp-content/uploads/2014/07/JungTMuscore;FinalTMuscore;Report2.pdf>.
- Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders*, 11, 33–47.
- Kadden, R. M., Cooney, N. L., Getter, H., & Litt, M. D. (1990). Matching alcoholics to coping skills or interactional therapies: Post-treatment results. *Journal of Consulting and Clinical Psychology*, 57, 698–704.
- Kagee, A., Nel, A., & Saal, W. (2014). Factor structure of the Beck Depression Inventory–II among South Africans receiving antiretroviral therapy. *AIDS Care*, 26(2), 257–262.
- Kahn, M. W., Fox, H., & Rhode, R. (1988). Detecting faking on the Rorschach: Computer versus expert clinical judgment. *Journal of Personality Assessment*, 52, 516–523.
- Kahn, R. L., & Cannell, C. F. (1961). *The dynamics of interviewing: Theory, technique, and cases*. New York, NY: Wiley.
- Kahn, T. C., & Giffen, M. B. (1960). *Psychological techniques in diagnosis and evaluation*. New York, NY: Pergamon Press.
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality*, 14(3), 265–278.
- Kamin, L. J. (1974). *The science and politics of I.Q.* Hillsdale, NJ: Erlbaum.
- Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice*, 31, 155–164.
- Kane, R. L. (2007). Editorial: Introduction to this supplement (Automated Neuropsychological Assessment Metric). *Archives of Clinical Neuropsychology*, 22, S3–S5.
- Kanfer, F. H., & Grimm, L. G. (1977). Behavioral analysis: Selecting target behaviors in the interview. *Behavior Modification*, 4, 419–444.
- Kanfer, F. H., & Saslow, G. (1969). Behavioral diagnosis. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status* (pp. 417–444). New York, NY: McGraw-Hill.
- Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *The WAIS-R as a neuropsychological instrument*. San Antonio, TX: Psychological Corporation.
- Kaplan, E., Fein, D., Morris, R., Kramer, J. H., & Delis, D. C. (1999). *The WISC-III as a process instrument*. San Antonio, TX: Psychological Corporation.
- Kaplan, R. M., & Saccuzzo, D. P. (1993). *Psychological testing: Principles, applications and issues* (3rd ed.). Belmont, CA: Wadsworth.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (5th ed.). Pacific Grove, CA: Brooks/Cole.
- Kareken, D. A., & Williams, J. M. (1994). Human judgment and estimation of premorbid intellectual function. *Psychological Assessment*, 6, 83–91.
- Karson, M. (2005). Overinterpretation of the Rorschach and the MMPI-2 when standard error is ignored. *Scientific Review of Mental Health Practice*, 3, 25–29.
- Kaslow, N. J., Bollini, Druss, B., Glueckauf, R. L., Goldfrank, R. R., Kelleher, K. J., ... Zeltzer, L. (2007). Health care for the whole person: Research update. *Professional Psychology: Research and Practice*, 38, 278–289.
- Katigbak, M. S., Church, A. T., & Akamine, T. X. (1996). Cross-cultural generalizability of personality dimensions: Relating indigenous and imported dimensions in two cultures. *Journal of Personality and Social Psychology*, 70(1), 99–114.

- Katon, W. J., & Walker, E. A. (1998). Medically unexplained symptoms in primary care. *Journal of Clinical Psychiatry*, 59, 15–21.
- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York, NY: Wiley.
- Kaufman, A. S. (2000). Seven questions about the WAIS-III regarding differences in abilities across the 16 to 89 year life span. *School Psychology Quarterly*, 15, 3–29.
- Kaufman, A. S., Kaufman, J. C., Balgopal, R., & McLean, J. E. (1996). Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology*, 25, 97–105.
- Kaufman, A. S., & Kaufman, N. L. (Eds.) (2001). *Specific learning disabilities and difficulties in children and adolescents: Psychological assessment and evaluation*. Cambridge, UK: Cambridge University Press.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition*. Minneapolis, MN: Pearson Assessments.
- Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York, NY: Wiley.
- Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York, NY: Wiley.
- Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence* (2nd ed.). Boston, MA: Allyn & Bacon.
- Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Boston, MA: Allyn & Bacon.
- Kaufman, A. S., Lichtenberger, E. O., & Fletcher-Janzen, E. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: Wiley.
- Kaufman, A. S., Lichtenberger, E. O., & McLean, J. E. (2001). Two- and three-factor solutions of the WAIS-III. *Assessment*, 8(3), 267–280.
- Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology*, 44, 231–248.
- Kay, R. (2007). A systematic evaluation of learning objects for secondary school students. *Journal of Educational Technology Systems*, 35(4), 411–448.
- Kearney, C. A., Cook, A. C., Chapman, C., & Bensaheb, A. (2006). Exploratory and factor analysis of the Motivation Assessment Scale and the Resident Choice Assessment Scale. *Journal of Developmental and Physical Disabilities*, 18, 1–11.
- Keiller, S., & Graham, J. R. (1993). The meaning of low scores on the MMPI-2 clinical scales of normal subjects. *Journal of Personality Assessment*, 61, 211–223.
- Keith, T. Z., Fine, J. D., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review*, 35, 108–127.
- Kelly, E. L., & Fiske, D. W. (1951). *The prediction of performance in clinical psychology*. Ann Arbor: University of Michigan Press.
- Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility with fidelity. *Professional Psychology: Research and Practice*, 38(1), 13–20.
- Kendall, P. C., & Hollon, S. D. (Eds.). (1981). *Assessment strategies for cognitive-behavioral interventions*. New York, NY: Academic Press.
- Kendall, P. C., & Treadwell, K. (2007). The role of self-statements as a mediator in treatment for anxiety-disordered youth. *Journal of Consulting and Clinical Psychology*, 75, 380–389.
- Kennedy, B. L., Schwab, J. J., Morris, R. L., & Beldia, G. (2001). Assessment of state and trait anxiety in subjects with anxiety and depressive disorders. *Psychiatric Quarterly*, 72(3), 263–276.
- Kerns, L. L. (1986). Falsifications in the psychiatric history: A differential diagnosis. *Psychiatry*, 49, 13–17.
- Killgore, W. D. S., & Dellapietra, L. (2000). Item response biases on the logical memory delayed recognition subtest of the Wechsler Memory Scale—III. *Psychological Reports*, 86, 851–857.

- Kim, B. S., Atkinson, D. R., & Yang, P. H. (1999). The Asian Values Scale: Development, factor analysis, validation, and reliability. *Journal of Counseling Psychology*, 46, 342–352.
- Kim, S., Goodman, G. M., Toruno, J. A., Sherry, A. R., & Kim, H. K. (2015). The cross-cultural validity of the MMPI-2–RF higher-order scales in a sample of North Korean female refugees. *Assessment*, 22(5), 640–649.
- Kinzel, R. L., & Murray, S. C. (1984). Kinesthesias in perception and the experience type: Dance and creative projection. *British Journal of Projective Psychology and Personality Study*, 29, 3–7.
- Kinder, B. N. (1994). Where the action is in personality assessment. *Journal of Personality Assessment*, 62, 585–588.
- King, L. C., Bailie, J. M., Kinney, D. I., & Nitch, S. R. (2012). Is the Repeatable Battery for the Assessment of Neuropsychological Status factor structure appropriate for inpatient psychiatry? An exploratory and higher-order analysis. *Archives of Clinical Neuropsychology*, 27(7), 756–765.
- Klassen, R. M. (2004). Optimism and realism: A review of self efficacy from a cross-cultural perspective. *International Journal of Psychology*, 39, 205–230.
- Kleespies, P. M. (2004). *Life and death decisions: Psychological and ethical considerations in end-of-life care*. Washington, DC: American Psychological Association.
- Klein, R. G. (1986). Questioning the usefulness of projective psychological tests for children. *Journal of Developmental and Behavioral Pediatrics*, 7, 378–382.
- Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, 107, 296–310.
- Klinefelter, D., Pancoast, D. L., Archer, R. P., & Pruitt, D. L. (1990). Recent adolescent MMPI norms: T score elevation comparisons to Marks and Briggs. *Journal of Personality Assessment*, 54, 379–389.
- Klopfer, B. (1937). The present status of the theoretical development of the Rorschach method. *Rorschach Research Exchange*, 1, 142–147.
- Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. (1956). *Developments in the Rorschach technique* (Vol. 2). Yonkers, NY: World Books.
- Klopfer, B., & Davidson, H. (1962). *The Rorschach technique: An introductory manual*. New York, NY: Harcourt, Brace, Jovanovich.
- Klopfer, W. G. (1960). *The psychological report*. New York, NY: Grune & Stratton.
- Knaster, C. A., & Micucci, J. A. (2013). The effect of client ethnicity on clinical interpretation of the MMPI-2. *Assessment*, 20(1), 42–47.
- Knight, R. G., & Godfrey, H. P. D. (1996). Psychosocial aspects of neurological disorders: Implications for research in neuropsychology. *Australian Psychologist*, 31, 48–51.
- Koenen, K. C., Stellman, J. M., Stellman, S. D., & Sommer, J. F., Jr. (2003). Risk factors for course of posttraumatic stress disorder among Vietnam veterans: A 14-year follow-up of American Legionnaires. *Journal of Consulting and Clinical Psychology*, 71(6), 980–986.
- Kohli, A., Rana, D. K., Gupta, N., & Kulhara, P. (2015). Neuropsychological assessment in obsessive-compulsive disorder. *Indian Journal of Psychological Medicine*, 37(2), 205–211.
- Kojima, M., Furukawa, T. A., Takahashi, H., Kawai, M., Nagaya, T., & Tokudome, S. (2002). Cross-cultural validation of the Beck Depression Inventory–II in Japan. *Psychiatry Research*, 110(3), 291–299.
- Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. W. (1999). Structure and diagnostic benefits of a normative taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology*, 37, 29–48.
- Koppitz, E. M. (1963). *The Bender Gestalt Test for Young Children* (Vol. 1). New York, NY: Grune & Stratton.
- Koppitz, E. M. (1975). *The Bender Gestalt Test for Young Children. Vol. 2: Research and applications 1963–1973*. New York, NY: Grune & Stratton.
- Kostlan, A. (1954). A method for the empirical study of psychodiagnosis. *Journal of Consulting Psychology*, 18, 83–88.

- Krishnamurthy, R., Archer, R. P., & Huddleston, E. N. (1995). Clinical research note on psychometric limitations of two Harris-Lingoes subscales for the MMPI-2. *Assessment*, 2(3), 301–304.
- Kramer, J. H., Kaplan, E., & Huckleba, W. (1999). Configural errors on WISC-III Block Design. *Journal of the International Neuropsychological Society*, 5, 518–524.
- Kratochwill, T. R. (1985). Selection of target behaviors in behavioral consultation. *Behavior Assessment*, 7, 49–61.
- Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research*, 21(3), 267–276.
- Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., ... Eisman, E. J. (2000). Empirical support for psychological assessment in clinical care settings. *Professional Psychology*, 31, 119–130.
- Kühn, R. (1963). Über die kritische Rorschach-Forschung und einige ihrer Ergebnisse. *Rorschachiana*, 8, 105–114.
- Kumar, G., Steer, R. A., Teitelman, K. B., & Villacis, L. (2002). Effectiveness of Beck Depression Inventory–II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment*, 9(2), 164–170.
- Kunce, J. T., & Tamkin, A. S. (1981). Rorschach movement and color responses and MMPI social extraversion and thinking introversion personality types. *Journal of Personality Assessment*, 45(1), 5–10.
- Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO–PI–R. *Journal of Personality Assessment*, 76(2), 315–332.
- Kutchins, H., & Kirk, S. A. (2003). *Making us crazy: DSM: The psychiatric bible and the creation of mental disorders*. New York, NY: Free Press.
- Kvaal, K., Ulstein, I., Nordhus, I. H., & Engedal, K. (2005). The Spielberger State-Trait Anxiety Inventory (STAI): The state scale in detecting mental disorders in geriatric patients. *International Journal of Geriatric Psychiatry*, 20(7), 629–634.
- Kvaal, S., Choca, J., & Groth-Marnat, G. (2003). The integrated psychological report. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrated assessment of adult personality* (2nd ed., pp. 398–434). New York, NY: Guilford Press.
- Kwan, K. K. (1999). MMPI and MMPI-2 performance of the Chinese: Cross-cultural applicability. *Professional Psychology: Research and Practice*, 30(3), 260–268.
- Lachar, D., & Gruber, C. P. (2001). *Personality Inventory for Children, Second Edition (PIC–2) standard form and behavioral summary manual*. Los Angeles, CA: Western Psychological Services.
- Lacks, P. (1984). *Bender-Gestalt screening for brain dysfunction*. New York, NY: Wiley.
- Lacks, P. (1999). *Bender-Gestalt screening for brain dysfunction* (2nd ed.). New York, NY: Wiley.
- Lacks, P. (2000). Visuoconstructive abilities. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 401–436). New York, NY: Wiley.
- Lacks, P., & Newport, K. (1980). A comparison of scoring systems and level of scorer experience on the Bender-Gestalt Test. *Journal of Personality Assessment*, 44, 351–357.
- Lamb, D. G., Berry, D. T., Wetter, M. W., & Baer, R. A. (1994). Effects of two types of information on malingering of closed head injury on the MMPI-2: An analog investigation. *Psychological Assessment*, 6, 8–13.
- Lambert, M. J., & Barley, D. E. (2002). Research summary on the therapeutic relationship and psychotherapy outcome. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 17–32). New York, NY: Oxford University Press.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., et al. (1996). *Administration and scoring manual for the Outcome Questionnaire (OQ45.2)*. Wilmington, DE: American Professional Credentialing Services.

- Lambert, M. J., & Hawkins, R. J. (2004). Measuring outcome in professional practice: Considerations in selecting brief outcome measures. *Professional Psychology: Research and Practice*, 35, 492–499.
- Landrine, H., & Klonoff, E. A. (1994). The African American Acculturation Scale: Development, reliability, and validity. *Journal of Black Psychology*, 20, 104–127.
- Lange, A. J., & Jakubowski, P. (1976). *Responsible assertive behavior: Cognitive/behavioral procedures for trainers*. Champaign, IL: Research Press.
- Lanyon, R. I., & Goodstein, L. D. (1982). *Personality assessment* (2nd ed.). New York, NY: Wiley.
- Lapouse, R., & Monk, M. A. (1958). An epidemiologic study of behavior characteristics of children. *American Journal of Public Health*, 48, 1134–1144.
- Lapouse, R., & Monk, M. A. (1964). Behavior deviations in a representative sample of children: Variations by sex, age, race, social class, and family size. *American Journal of Orthopsychiatry*, 34, 436–446.
- Larkin, K. T. (2006). Psychophysiological assessment. In M. Hersen (Ed.), *Clinician's handbook of behavioral assessment* (pp. 165–185). New York, NY: Elsevier.
- Larrabee, G. J. (2005). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach*. (pp. 115–158). New York, NY: Oxford University Press.
- Larrabee, G. J. (1986). Another look at VIQ–PIQ scores and unilateral brain damage. *International Journal of Neuroscience*, 29, 141–148.
- Larson, E. B., Kirschner, K., Bode, R., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the Repeatable Battery for the Assessment of Neuropsychological Status in the evaluation of stroke patients. *Journal of Clinical and Experimental Neuropsychology*, 27, 16–32.
- La Rue, A., & Jarvik, L. R. (1987). Cognitive function and prediction of dementia in old age. *International Journal of Aging and Human Development*, 25, 78–89.
- Laurent, H., Chinot, L., Laure, J., Plancherel, B., Sofia, C., Halfon, O., & Randolph, C. (2007). Detection of cognitive impairment with the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in adolescents with psychotic symptomology. *Schizophrenia Research*, 95, 48–53.
- Lawrence, S. B. (1984). *Lawrence Psychological–Forensic Examination (Law–PSI)*. San Bernadino, CA: Lawrence Psychological Center.
- Lazarus, A. A. (1973). Multimodel behavior therapy: Treating the “BASIC ID.” *Journal of Nervous and Mental Disease*, 156, 404–411.
- Lazarus, A. A. (1989). *The practice of multi-modal therapy*. Baltimore, MD: Johns Hopkins University Press.
- Lazarus, A. A. (2005). Multimodal therapy. In R. J. Corsini & D. Wedding (Eds.) *Current psychotherapies* (pp. 737–371). Belmont, CA: Thomson Brooks/Cole.
- Leary, T. (1957). *Interpersonal diagnosis of personality*. New York, NY: Ronald Press.
- Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, 68, 203–210.
- Lefebvre, M. F. (1981). Cognitive distortions and cognitive errors in depressed and low back pain patients. *Journal of Consulting and Clinical Psychology*, 49, 517–525.
- Leffard, S. A., Miller, J. A., Bernstein, J., DeMann, J. J., Mangis, H. A., & McCoy, E. L. B. (2006). Substantive validity of working memory measures in major cognitive functioning test batteries for children. *Applied Neuropsychology*, 13, 230–241.
- Lefkowitz, J., & Fraser, A. W. (1980). Assessment of achievement and power motivation of blacks and whites, using a black and white TAT with black and white administrators. *Journal of Applied Psychology*, 65, 685–696.
- Lehne, G. K. (2002). The NEO Personality Inventory and the Millon Clinical Multiaxial Inventory in the forensic evaluation of sex offenders. In P. T. Costa Jr., & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (2nd ed., pp. 269–282). Washington, DC: American Psychological Association.

- Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review, 21*, 401–419.
- Leigh, J., & Zaylor, C. (2000). Cyberspace: Creating a therapeutic environment for telehealth applications. *Professional Psychology: Research and Practice, 31*, 478–483.
- Lemsky, C. (2000). Neuropsychological assessment and treatment planning. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 535–576). New York, NY: Wiley.
- Lenzenweger, M. F. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry, 56*, 1009–1015.
- Lessenger, L. H. (1997). Acculturation and MMPI-2 scale scores of Mexican American substance abuse patients. *Psychological Reports, 80*(3), 1181–1182.
- Leura, A. V., & Exner, J. E. (1978). *Structural differences in the records of adolescents as a function of being tested by one's own teacher*. Rorschach Workshops (Study No. 265, unpublished).
- Levak, R. W., Siegel, L., & Nichols, D. S. (2011). *Therapeutic feedback with the MMPI-2: A positive psychology approach*. New York, NY: Routledge.
- Levenson, M. R., Aldwin, C. M., Bossé, R., & Spiro, A. (1988). Emotionality and mental health: Longitudinal findings from the normative aging study. *Journal of Abnormal Psychology, 97*(1), 94–96.
- Levin, H. S., High, W. M., Goethe, K. E., Sisson, R. A., Overall, J. E., Rhoades, H. M., ... & Gary, H. E. (1987). The Neurobehavioral Rating Scale: Assessment of the behavioral sequelae of head injury by the clinician. *Journal of Neurology, Neurosurgery, and Psychiatry, 50*, 183–193.
- Levine, A. (2007). Cultural beliefs keep many Hispanics away from getting mental health care. *Psychiatric News, 42*, 8–9.
- Levine, D. (1981). Why and when to test: The social context of psychological testing. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 553–580). New York, NY: Springer.
- Levitt, E. E., & Gotts, E. E. (1995). *The clinical application of MMPI special scales* (2nd ed.). New York, NY: Routledge.
- Lewak, R., & Hogan, L. (2003). Applying assessment information. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (2nd ed., pp. 356–397). New York, NY: Guilford Press.
- Lewinsohn, P. M., Rohde, P., Klein, D. N., & Seeley, J. R. (1999). Natural course of adolescent major depressive disorder. I: Continuity into young adulthood. *Journal of the American Academy of Child Adolescent Psychiatry, 38*, 56–63.
- Lezak, M. D. (1988). IQ: RIP. *Journal of Experimental and Clinical Neuropsychology, 10*, 351–361.
- Lezak, M. D. (1989). Assessment of psychosocial dysfunctions resulting from head trauma. In M. D. Lezak (Ed.), *Assessment of the behavioral consequences of head trauma: Frontiers of clinical neuroscience* (Vol. 7). New York, NY: Alan R. Liss.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press.
- Lichtenberger, E. O. (2006). Computer utilization and clinical judgment in psychological assessment reports. *Journal of Clinical Psychology, 62*, 19–32.
- Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.
- Lichtenberger, E. O., Kaufman, A. S., & Lai, Z. C. (2002). *Essentials of WMS-III assessment*. New York, NY: Wiley.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159–183.
- Lightfoot, S. L., & Oliver, J. M. (1985). The Beck Inventory: Psychometric properties in university students. *Journal of Personality Assessment, 49*, 434–436.
- Lilienfeld, S. O., Wood, J. W., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*(2), 27–66.

- Lim, C., Sim, K., Renjan, V., Sam, H. F., & Quah, S. L. (2014). Adapted cognitive-behavioral therapy for religious individuals with mental disorder: A systematic review. *Asian Journal of Psychiatry*, 9, 3–12.
- Lin, E. J., & Church, A. T. (2004). Are indigenous Chinese personality dimensions culture-specific? An investigation of the Chinese Personality Assessment Inventory in Chinese American and European American samples. *Journal of Cross-Cultural Psychology*, 35(5), 586–605.
- Lindsay, D. S., & Read, J. D. (1995). “Memory work” and recovered memories of childhood sexual abuse: Scientific evidence and public and professional issues. *Psychology, Public Policy, and Law*, 1, 846–908.
- Lindsay, K. A., Sankis, L. M., & Widiger, T. A. (2000). Gender bias in self-report personality disorder inventories. *Journal of Personality Disorders*, 14(3), 218–232.
- Linger, M. L., Ray, G. E., Zachar, P., Underhill, A. T., & LoBello, S. G. (2007). Decreasing scoring errors on Wechsler scale Vocabulary, Comprehension, and Similarities subtests: A preliminary study. *Psychological Reports*, 101, 661–669.
- Linscott, J., & DiGiuseppe, R. (1998). Cognitive assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 104–125). Boston, MA: Allyn & Bacon.
- Little, K. B., & Shneidman, E. S. (1959). Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs*, 73 (6, Whole No. 476).
- Lloyd, A. M. (2009). *The impact of culture on the MCMI-III scores of African American and Caribbean Blacks*. (Unpublished doctoral dissertation). Nova Southeastern University, Ft. Lauderdale, FL.
- Lodge, J., Tripp, G., & Harte, D. K. (2000). Think aloud, thought listing, and video-mediated recall procedures in the assessment of children’s self talk. *Cognitive Therapy and Research*, 24, 399–418.
- Loe, S. A., Kadlubek, R. M., & Marks, W. (2007). Administration and scoring errors for the WISC-IV among graduate student examiners. *Journal of Psychoeducational Assessment*, 25, 237–247.
- Loenberger, L. T. (1989). The question of organicity: Is it still functional? *Professional Psychology: Research and Practice*, 20, 411–414.
- Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist*, 48, 518–537.
- Long, J. D., Harring, J. R., Brekke, J. S., Test, M. A., & Greenberg, J. (2007). Longitudinal construct validity of Brief Symptom Inventory Subscales in schizophrenia. *Psychological Assessment*, 19, 298–308.
- Loranger, A. W. (1988). *Personality Disorder Examination (PDE) manual*. Yonkers, NY: DV Communications.
- Loranger, A. W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S. M., ... Regier, D. A. (1995). The International Personality Disorder Examination. *Archives of General Psychiatry*, 51, 215–224.
- Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: A meta-analysis. *Psychological Bulletin*, 130, 531–552.
- Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1989). Verbal and visual memory index discrepancies from the Wechsler Memory Scale-Revised: Cautions in interpretation. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3, 198–202.
- Lovibund, P. F., & Lovibund, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scale (DASS) with the Beck Depression Inventories. *Behaviour Research and Therapy*, 33, 335–342.
- Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935–1982. *American Psychologist*, 39, 451–454.
- Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1985). Psychological test usage patterns in five professional settings. *American Psychologist*, 40, 857–861.

- Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1986). Selected characteristics of psychologists and psychological assessment in five settings: 1959–1988. *Professional Psychology: Research and Practice*, 17, 155–157.
- Luborsky, L. (1994). Therapeutic alliances as predictors of psychotherapy outcomes: Factors explaining predictive success. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 38–50). New York, NY: Wiley.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies. *Archives of General Psychiatry*, 32, 995–1008.
- Lucas, C. P., Zhang, H., Fisher, P. W., Shaffer, D., Regier, D. A., Narow, W. E., ... Friman, P. (2001). The DISC Predictive Scales (DPS): Efficiently screening for diagnosis. *Journal of the Academy of Child and Adolescent Psychiatry*, 40, 443–449.
- Luciana, M. (2003). Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological testing Automated Battery (CANTB). *Journal of Child Psychology and Psychiatry*, 44, 649–663.
- Luria, A. R. (1973). *The working brain*. New York, NY: Basic Books.
- Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York, NY: Basic Books.
- Lyle, O. E., & Gottesman, I. I. (1977). Premorbid psychometric indicators of the gene for Huntington's disease. *Journal of Consulting and Clinical Psychology*, 45(6), 1011–1022.
- Lynn, R. (1977). The intelligence of the Japanese. *Bulletin of the British Psychological Society*, 30, 69–72.
- Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly*, 12, 197–234.
- Maddox, T. (2003). *Tests: A comprehensive reference for assessments in psychology, education, and business* (5th ed.). Austin, TX: PRO-ED.
- Magnavita, J. J. (2008). Using the MCMI in general treatment planning. In T. Millon & C. Bloom (Eds.), *The Millon Inventories: A practitioner's guide to personalized clinical assessment* (2nd ed. pp. 327–346). New York, NY: Guilford Press.
- Makatura, T. J., Lam, C. S., Leahy, B. J., Castillo, M. T., & Kalpakjian, C. Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury*, 13, 355–367.
- Mallinckrodt, B. (1996). Change in working alliance, social support, and psychological symptoms in brief therapy. *Journal of Counseling Psychology*, 43, 448–455.
- Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. New York, NY: Oxford University Press.
- Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology*, 626(1), 83–86.
- Marchese, M. C., & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment*, 1, 18–26.
- Marek, R. J., Ben-Porath, Y. S., Windover, A., Tarescavage, A. M., Merrell, J. Ashton, K., ... Heinberg, L. J. (2013). Assessing psychosocial functioning of bariatric surgery candidates with the Minnesota Multiphasic Personality Inventory–2 Restructured Form (MMPI-2-RF). *Obesity Surgery*, 23(11), 1864–1873.
- Marek, R. J., Block, A. R., & Ben-Porath, Y. S. (2015). The Minnesota Multiphasic Personality Inventory–2–Restructured Form (MMPI-2-RF): Incremental validity in predicting early postoperative outcomes in spine surgery candidates. *Psychological Assessment*, 27(1), 114–124.
- Margolin, G., Hattem, D., John, R. S., & Yost, K. (1985). Perceptual agreement between spouses and outside observers when coding themselves and a stranger dyad. *Behavioral Assessment*, 7, 235–247.
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88(1), 139–157.

- Marks, I. M., Cavanagh, K., & Gega, L. (2007). *Hands-on help: Computer aided psychotherapy*. London, UK: Taylor & Francis.
- Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults*. New York, NY: Oxford University Press.
- Marley, M. L. (1982). *Organic brain pathology and the Bender Gestalt Test: A differential diagnostic scoring system*. New York, NY: Grune & Stratton.
- Martin, P., Churchard, M., Kotcher, S., & Korenblum, M. (1991). Diagnostic utility of the Beck Depression Inventory with adolescent psychiatric outpatients and inpatients. *Canadian Journal of Psychiatry*, 36, 428–431.
- Martinez, S., Stillerman, L., & Waldo, M. (2005). Reliability and validity of the SCL-90-R with Hispanic college students. *Hispanic Journal of Behavioral Sciences*, 27, 254–264.
- Maruish, M. E. (Ed.). (2000). *Handbook of psychological testing in primary care settings*. Mahwah, NJ: Erlbaum.
- Maruish, M. E. (2004). *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed.). Mahwah, NJ: Erlbaum.
- Maruish, M. E., & Nelson, E. A. (2014). *Psychological testing in the age of managed behavioral health care*. Mahwah, NJ: Erlbaum.
- Masling, J. (1992). The influence of situation and interpersonal variables in projective testing. *Journal of Personality Assessment*, 59, 616–640.
- Masling, M. (2006). When Homer nods: An examination of some of the systematic errors in Rorschach scholarship. *Journal of Personality Assessment*, 87, 62–73.
- Matarazzo, J. D. (1965). The interview. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 403–450). New York, NY: McGraw-Hill.
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist*, 45, 999–1017.
- Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *Journal of Clinical Psychology*, 44, 940–950.
- Matarazzo, J. D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample. *Psychological Assessment*, 1, 186–191.
- Mauger, P. A. (1972). The test-retest reliability of persons: An empirical investigation utilizing the MMPI and the personality research form. *Dissertation Abstracts International*, 33, 2816B.
- Mayes, S. D., & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children—Third and Fourth Edition predictors of academic achievement in children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly*, 22, 234–249.
- Mayes, S. D., Calhoun, S. L., & Crowell, E. W. (1998). WISC-III profiles for children with and without learning disabilities. *Psychology in the Schools*, 35, 309–316.
- Maziade, M., Roy, A. A., Fournier, J. P., Cliche, D., Merette, C., Caron, C., ... & Raymond, V. (1992). Reliability of best-estimate diagnosis in genetic linkage studies of major psychoses. *American Journal of Psychiatry*, 149, 1674–1686.
- McCall, R. B., Appelbaum, M. I., & Hogarty, P. S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development*, 38(3, Serial No. 150), 1–83.
- McCallum, R. S., Bracken, B. A., & Wasserman, J. (2001). *Essentials of nonverbal assessment*. New York, NY: Wiley.
- McCann, J. T. (1991). Convergent and discriminant validity of the MCMI-II and MMPI personality disorder scales. *Psychological Assessment*, 3, 9–18.
- McCloskey, G., & Maerlander, A. (2005). The WISC-IV integrated. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist practitioner perspectives* (pp. 101–149). San Diego, CA: Elsevier.
- McConaughy, E. A., DiClemente, C. C., Prochaska, J. O., & Velicer, W. F. (1989). Stages of change in psychotherapy: A follow-up report. *Psychotherapy: Theory, Research, Practice, Training*, 26(4), 494–503.

- McConaughy, E. A., Prochaska, J. O., & Velicer, W. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, Research, and Practice*, 20, 368–375.
- McCormick, I. A. (1984). A simple version of the Rathus Assertiveness Schedule. *Behavioral Assessment*, 7, 95–99.
- McCown, W., Fink, A. D., Galina, H., & Johnson, J. (1992). Effects of laboratory-induced controllable and uncontrollable stress on Rorschach variables m and Y. *Journal of Personality Assessment*, 59, 564–573.
- McCrae, R. R. (2002). NEO-PI-R data from 36 cultures. In J. Allik & R. R. McCrae (Eds.), *The five-factor model of personality across cultures* (pp. 105–125). New York, NY: Kluwer Academic/Plenum.
- McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, 90(2), 105–109.
- McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer Academic/Plenum.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90.
- McCrae, R. R., & Costa, P. T., Jr. (1992). Discriminant validity of NEO-PI-R facet scales. *Educational and Psychological Measurement*, 52, 229–237.
- McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Personality theories and models* (Vol. 1, pp. 273–294). London, UK: Sage.
- McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories: Professional manual*. Lutz, FL: Psychological Assessment Resources.
- McCrae, R. R., Costa, P. T., Jr., de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., ... Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology*, 35(2), 466–477.
- McCrae, R. R., Costa, P. T., Jr., Del Pilar, G. H., Rolland, J., & Parker, W. D. (1998). Cross-cultural assessment of the five-factor model: The revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 29(1), 171–188.
- McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment*, 84(3), 261–270.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15(1), 28–50.
- McCrae, R. R., Martin, T. A., & Costa, P. T., Jr. (2005). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment*, 12(4), 363–373.
- McCrae, R. R., Terracciano, A., & Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88(3), 547–561.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, M. (1992). Illusions of meaning in the ipsative assessment of children's abilities. *Journal of Special Education*, 25, 504–526.
- McFall, R. M., & Lillesand, D. V. (1971). Behavior rehearsal with modeling and coaching in assertive training. *Journal of Abnormal Psychology*, 77, 313–323.
- McFie, J. (1960). Psychological testing in clinical neurology. *Journal of Nervous and Mental Disease*, 131, 383–393.
- McFie, J. (1969). The diagnostic significance of disorders of higher nervous activity syndromes related to frontal, temporal, parietal, and occipital lesions. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol. 4, pp. 1–12). New York, NY: American Elsevier.

- McGlyn, F. D., & Rose, M. P. (1998). Assessment of anxiety and fear. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 179–209). Boston, MA: Allyn & Bacon.
- McGrath, R. E. (2001). Toward more clinically relevant research. *Journal of Personality Assessment*, 77, 307–332.
- McGrath, R. E., & Ingersoll, J. (1999a). Writing a good cookbook. I: A review of MMPI high-point code system studies. *Journal of Personality Assessment*, 73, 149–178.
- McGrath, R. E., & Ingersoll, J. (1999b). Writing a good cookbook. II: A synthesis of MMPI high-point code system study effect sizes. *Journal of Personality Assessment*, 73, 179–198.
- McGrath, R. E., Powis, D., & Pogge, D. L. (1998). Code type-specific tables for interpretation of MMPI-2 Harris and Lingoes Subscales: Consideration of gender and code definition. *Journal of Clinical Psychology*, 54, 655–664.
- McIntosh, J. A., Belter, R. W., Saylor, C. F., Finch, A. J., & Edwards, G. L. (1988). The Bender-Gestalt with adolescents: Comparison of two scoring systems. *Journal of Clinical Psychology*, 44, 226–230.
- McKay, C., Wertheimer, J. C., Fichtenberg, N. L., & Casey, J. E. (2007). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Clinical utility in a traumatic brain injury sample. *Clinical Neuropsychology*, 28, 1–14.
- McLean, P., & Taylor, S. (1992). Severity of unipolar depression and choice of treatment. *Behavior Research and Therapy*, 30, 443–451.
- McLeod, B. D., Jensen-Doss, A., & Ollendick, T. H. (Eds.). (2013). *Diagnostic and behavioral assessment in children and adolescents: A clinical guide*. New York, NY: Guilford Press.
- McLeod, C. C., Budd, M. A., & McClelland, D. C. (1997). Treatment of somatization in primary care. *General Hospital Psychiatry*, 19, 251–258.
- McMinn, M. R., Bearse, J., Heyne, L. K., Smithberger, A., & Erb, A. L. (2011). Technology and independent practice: Survey findings and implications. *Professional Psychology: Research and Practice*, 42(2), 176–184.
- McMinn, M. R., Buchanan, T., Ellens, B. M., & Ryan, M. K. (1999). Technology, professional practice, and ethics: Survey findings and implications. *Professional Psychology: Research and Practice*, 30, 165–172.
- McMinn, M. R., Ellens, B. M., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretation. *Assessment*, 6, 71–77.
- McNulty, J. L., Ben-Porath, Y. S., & Graham, J. R. (1998). An empirical examination of the correlates of well-defined and not-defined MMPI-2 code types. *Journal of Personality Assessment*, 71, 393–410.
- McNulty, J. L., Graham, J. R., Ben-Porath, Y. S., & Stein, L. A. R. (1997). Comparative validity of MMPI-2 scores of African American and Caucasian mental health center clients. *Psychological Assessment*, 9, 464–470.
- McShane, D. A., & Plas, J. M. (1984). The cognitive functioning of American Indian children: Moving from the WISC to the WISC-R. *School Psychology Review*, 13, 16–73.

- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality*, 1, 27–32.
- Megargee, E. I. (1972). *The California Psychological Inventory handbook*. San Francisco, CA: Jossey-Bass.
- Megargee, E. I., & Mendelsohn, G. A. (1962). A cross-validation of twelve MMPI indices of hostility and control. *Journal of Abnormal Psychology*, 65, 431–438.
- Mehrabian, A. (1972). *Nonverbal communication*. Chicago, IL: Aldine-Atherton.
- Melzack, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, 1, 277–299.
- Mendoza, J. (2001). *Reporting the results of the neuropsychological evaluation*. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 95–122). Lutz, FL: Psychological Assessment Resources.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Psychological Assessment*, 7, 741–749.
- Meyer, G. J. (1997). On the integration of personality assessment methods: The Rorschach and the MMPI. *Journal of Personality Assessment*, 68, 297–330.
- Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Research*, 75, 46–81.
- Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Research and Practice*, 8, 389–386.
- Meyer, G. J. (2002). Exploring possible ethnic differences and bias in the Rorschach Comprehensive System. *Journal of Personality Assessment*, 78, 104–129.
- Meyer, G. J. (2004). The reliability of the Rorschach and Thematic Apperception Test (TAT) compared to other psychological and medical procedures: An analysis of systematically gathered evidence. In M. J. Hilsenroth & D. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 315–342). Hoboken, NJ: Wiley.
- Meyer, G. J. (Ed.) (2007). The MMPI-2 restructured clinical scales [Special issue]. *Journal of Personality Assessment*, 87(2).
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment*, 13, 486–502.
- Meyer, G. J., Erdberg, P., & Shaffer, T. W. (2007). Towards international normative reference data for the Comprehensive System [Special issue]. *Journal of Personality Assessment*, 89, S201–S216.
- Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., ... Reed, G. M. (2001). Psychological testing and psychological assessment. *American Psychologist*, 56, 128–165.
- Meyer, G. J., Giromini, L., Viglione, D. J., Reese, J. B., & Mihura, J. L. (2015). The association of gender, ethnicity, age, and education with Rorschach scores. *Assessment*, 22(1), 46–64.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69, 1–38.
- Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C., & Resnick, J. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment*, 78, 219–274.
- Meyer, G. J., Mihura, J., & Smith, B. (2005). The interclinician reliability of Rorschach interpretation in four data sets. *Journal of Personality Assessment*, 84, 296–314.
- Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual*. Toledo, OH: Rorschach Performance Assessment System.
- Meyers, J. E., & Meyers, K. R. (1996). *Rey Complex Figure Test and recognition trial: A professional manual*. Odessa, FL: Psychological Assessment Resources.

- Michaels, M. (2007). *Professional guide to psychological report writing*. New York, NY: Oxford University Press.
- Mihura, J. L., Meyer, G. J., Bombel, G., & Dumatriscu, N. (2015). Standards, accuracy, and questions of bias in Rorschach meta-analyses: Reply to Wood, Garb, Nezworski, Lilienfeld, and Duke (2015). *Psychological Bulletin*, *141*(1), 250–260.
- Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin*, *139*(3), 548–605.
- Millberg, W. P., Hebben, N., & Kaplan, E. (1996). The Boston Process Approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed., pp. 58–80). New York, NY: Oxford University Press.
- Miller, D. N. (2007). Projective techniques and the school-based assessment of childhood internalizing disorders: A critical analysis. *Journal of Projective Psychology and Mental Health*, *14*, 48–58.
- Miller, H. R., Streiner, D. L., & Parkinson, A. (1992). Maximum likelihood estimates of the ability of the MMPI and the MCMI personality disorder scales and the SIDP to identify personality disorders. *Journal of Personality Assessment*, *59*, 1–13.
- Miller, J. B., Axelrod, B. N., Rapport, L. J., Millis, S. R., VanDyke, S., Schutte, C., & Hanks, R. A. (2012). Parsimonious prediction of Wechsler Memory Scale, Fourth Edition scores: Immediate and delayed memory indexes. *Journal of Clinical and Experimental Neuropsychology*, *34*(5), 531–542.
- Miller, L. (1993, January/February). Toxic torts: Clinical neuropsychological and forensic aspects of chemical and electrical injuries. *Journal of Cognitive Rehabilitation*, *11*, 6–18.
- Miller, W. R., & Tonigan, J. S. (1996). Assessing drinkers' motivation for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*, *10*(2), 81–89.
- Miller, T. (1991). The psychotherapeutic utility of the five-factor model of personality: A clinician's experience. *Journal of Personality Assessment*, *57*(3), 415–433.
- Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale–III. *Journal of Clinical and Experimental Neuropsychology*, *21*, 87–93.
- Millon, T. (1969). *Modern psychopathology: A biosocial approach to maladaptive learning and functioning*. Philadelphia, PA: Saunders.
- Millon, T. (1977). *Millon Clinical Multiaxial Inventory*. Minneapolis, MN: National Computer Systems.
- Millon, T. (1987). *Manual for the MCMI-II* (2nd ed.). Minneapolis, MN: National Computer Systems.
- Millon, T. (1992). Millon Clinical Multiaxial Inventory: I & II. *Journal of Counseling and Development*, *70*, 422–426.
- Millon, T. (1993). *The MACI manual*. Minneapolis, MN: National Computer Systems.
- Millon, T. (1994). *Manual for the MCMI-III*. Minneapolis, MN: National Computer Systems.
- Millon, T. (1997). *Millon Clinical Multiaxial Inventory–III manual* (2nd ed.). Minneapolis, MN: National Computer Systems.
- Millon, T. (1999). *Personality guided therapy*. New York, NY: Wiley.
- Millon, T. (2011). *Disorders of personality: Introducing a DSM/ICD spectrum from normal to abnormal*. Hoboken, NJ: Wiley.
- Millon, T., & Bloom, C. (2008). *The Millon inventories: A practitioner's guide to personalized clinical assessment* (2nd ed.). New York, NY: Guilford Press.
- Millon, T., & Davis, R. D. (1996). *Disorders of personality: DSM-IV and beyond* (2nd ed.). New York, NY: Wiley.
- Millon, T., Green, C. J., & Meagher, R. B. (2000). *Manual for the Millon Behavioral Health Inventory*. Minneapolis, MN: National Computer Systems.
- Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory–IV (MCMI-IV) manual*. Minneapolis, MN: Pearson Clinical Assessments.

- Millon, T., Millon, C., Davis, R., & Grossman, S. (2006). *MCMI-III Manual* (3rd ed.). Minneapolis, MN: NCS Pearson.
- Millon, T., Millon, C., Meagher, S., Grossman, S., & Ramnath, R. (2004). *Personality disorders in modern life* (2nd ed.). Hoboken, NJ: Wiley.
- Milne, D. (1984). Improving the social validity and implementation of behavior therapy training for psychiatric nurses using a patient-centered learning format. *British Journal of Clinical Psychology*, 23, 313–314.
- Miranda, J., Azocar, F., Organista, K. C., Dwyer, E., & Areane, P. (2003). Treatment of depression among impoverished primary care patients from ethnic minority groups. *Psychiatric Services*, 54(2), 219–225.
- Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.
- Mitrushina, M., Boone, K., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York, NY: Oxford University Press.
- Mittenberg, W., Hammeke, T. A., & Rao, S. M. (1989). Intrasubtest scatter on the WAIS-R as a pathognomonic sign of brain injury. *Psychological Assessment*, 1, 273–276.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- Mizes, J. S., & Christiano, B. A. (1994). Assessment of cognitive variables relevant to cognitive behavioral perspectives on anorexia nervosa and bulimia nervosa. *Behavior Research and Therapy*, 33, 95–105.
- Mohr, D. C. (1995). The role of proscription in psychotherapy. *Psychotherapy*, 32, 187–193.
- Molish, H. B. (1967). Critique and problems of the Rorschach: A survey. In S. J. Beck & H. B. Molish (Eds.), *Rorschach's test. Vol. 2: A variety of personality pictures* (2nd ed., pp. 45–48). New York, NY: Grune & Stratton.
- Monahan, J., & Steadman, H. J. (2001). Violence risk assessment: A quarter century of research. In L. E. Frost & R. J. Bonnie (Eds.), *The evolution of mental health law* (pp. 195–211). Washington, DC: American Psychological Association.
- Monahan, J., Steadman, H. J., Robbins, P. C., Silver, E., Applebaum, P. S., Grisso, E. P., ... & Roth, L. H. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry*, 176(4), 312–319.
- Moon, G. W., Blakey, W. A., Gorsuch, R. L., & Fantuzzo, J. W. (1991). Frequent WAIS-R administration errors: An ignored source of inaccurate measurement. *Professional Psychology*, 22, 256–258.
- Moon, G. W., Fantuzzo, J. W., & Gorsuch, R. L. (1986). Teaching WAIS-R administration skills: Comparison of the MASTERY model to other existing clinical training modalities. *Professional Psychology*, 17, 31–35.
- Moos, R. H. (1990). Depressed patients' life context, amount of treatment, and treatment outcome. *Journal of Nervous and Mental Disease*, 178, 105–112.
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist*, 50(1), 14–23.
- Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (1998). Teaching and learning the Personality Assessment Inventory (PAI). In L. Handler & M. Hilsenroth (Eds.), *Teaching and learning psychological assessment* (pp. 191–214). Mahwah, NJ: Erlbaum.
- Morey, L. C. (2003). *Essentials of PAI Assessment*. Hoboken, NJ: Wiley.
- Morey, L. C. (2007). *Personality Assessment Inventory professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Morey, L. C., Blashfield, R. K., Webb, W. W., & Jewell, J. (1988). MMPI scales for DSM-III personality disorders: A preliminary validation study. *Journal of Clinical Psychology*, 44, 47–50.

- Morey, L. C., Gunderson, J. G., Quigley, B. D., Shea, M. T., Skodol, A. E., McGlashan, T. H., ... & Zanarini, M. C. (2002). The representation of borderline, avoidant, obsessive-compulsive, and schizotypal personality disorders by the five-factor model. *Journal of Personality Disorders*, 16(3), 215–234.
- Morey, L. C., Quigley, B. D., Sanislow, C. A., Skodol, A. E., McGlashan, T. H., Shea, M. T., ... Gunderson, J. G. (2002). Substance or style? An investigation of the NEO-PI-R validity scales. *Journal of Personality Assessment*, 79(3), 583–599.
- Morgan, C. D., Schoenberg, M. R., Dorr, D., & Burke, M. J. (2002). Overreport on the MCMI-III: Concurrent validation with the MMPI-2 using a psychiatric inpatient sample. *Journal of Personality Assessment*, 78, 288–300.
- Morgan, G., Gliner, J. A., & Harmon, R. J. (2001). Measurement validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 729–731.
- Morlan, K. K., & Tan, S. Y. (1998). Comparison of the Brief Psychiatric Rating Scale and the Brief Symptom Inventory. *Journal of Clinical Psychology*, 54, 885–894.
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD). Part 1: Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159–1165.
- Morrow, L. A., Furman, J. M. R., Ryan, C. M., & Hodgson, M. J. (1988). Neuropsychological deficits associated with verbatim abnormalities in solvent exposed workers. *Clinical Neuropsychologist*, 2, 272–273.
- Moser, R. S., & Schatz, P. (2002). Enduring effects of concussion in youth athletes. *Archives of Clinical Neuropsychology*, 17, 91–100.
- Moser, R. S., Schatz, P., & Jordan, B. D. (2005). Prolonged effects of concussion in high school athletes. *Neurosurgery*, 57, 300–306.
- Mostofsky, D., & Barlow, D. H. (2000). *The management of stress and anxiety in medical disorders*. Needham Heights, MA: Allyn & Bacon.
- Mueller, H. H., Dash, V. N., Matheson, D. W., & Short, R. H. (1984). WISC-R subtest patterning of below average, average, and above average I.Q. children: A meta-analysis. *Alberta Journal of Educational Research*, 30, 68–85.
- Mulvey, E., & Cauffman, E. (2001). The inherent limits of predicting school violence. *American Psychologist*, 56, 797–802.
- Mundia, L. (2011). Social desirability, non-response bias and reliability in a long self-report measure: Illustrations from the MMPI-2 administered to Brunei student teachers. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 31(2), 207–224.
- Munley, P. H. (1991). Confidence intervals for the MMPI-2. *Journal of Personality Assessment*, 57, 52–60.
- Murphy, J. M., Monson, R. R., Laird, N. M., Sobol, A. M., & Leighton, A. H. (2000). A comparison of diagnostic interviews for depression in the Stirling County Study: Challenges for psychiatric epidemiology. *Archives of General Psychiatry*, 57, 230–236.
- Murphy, L. L., Geisinger, K. F., Carlson, J. F., & Spies, R. A. (Eds.). (2011). *Tests in print VIII*. Lincoln, NE: Buros Institute of Mental Measurements.
- Murphy, M. J., Levant, R. G., Hall, J. E., & Glueckauf, R. L. (2007). Distance education in professional training in psychology. *Professional Psychology: Research and Practice*, 38, 97–103.
- Nathan, P. E., & Gorman, J. M. (Eds.). (2015). *A guide to treatments that work* (4th ed.). New York, NY: Oxford University Press.
- Neighbors, H. W., Caldwell, C., Williams, D. R., Nesse, R., Taylor, R. J., Bullard, K. M., ... Jackson, J. S. (2007). Race, ethnicity, and the use of services for mental disorders: Results from the National Survey of American Life. *Archives of General Psychiatry*, 64(4), 485–494.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.

- Nelson, H. E., & Williams, J. R. (1991). *National Adult Reading Test (NART): Test manual*. Windsor, England: NFER.
- Nelson, J. R., Roberts, M. L., Rutherford, R. B., Mathur, S. R., & Aaroe, L. A. (1999). A statewide survey of special education administrators and school psychologists regarding functional behavioral assessment. *Education and Treatment of Children*, 22, 267–279.
- Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *Clinical Neuropsychologist*, 20, 39–58.
- Nelson, R. E., & Maser, J. D. (1988). The DSM-III and depression: Potential contributions of behavioral assessment. *Behavioral Assessment*, 10, 45–66.
- Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment*, 15(4), 521–531.
- Nelson-Gray, R. O., & Farmer, R. F. (1999). Behavioral assessment of personality disorders. *Behavior Research and Therapy*, 37, 347–368.
- Nelson-Gray, R. O., & Paulson, J. (2004). Behavioral assessment and the DSM system. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3, pp. 470–488). Hoboken, NJ: Wiley.
- Newham, J. J., Westwood, M., Aplin, J. D., & Wittkowski, A. (2012). State-Trait Anxiety Inventory (STAI) scores during pregnancy following intervention with complementary therapies. *Journal of Affective Disorders*, 142(1), 22–30.
- Newman, F. L., Ciarlo, J. A., & Carpenter, D. (1999). Guidelines for selecting psychological instruments for treatment planning and outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 153–170). Mahwah, NJ: Erlbaum.
- Nezu, A. M., Nezu, C. M., Peacock, M. A., & Girdwood, C. P. (2004). Case formulation in cognitive-behavior therapy. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3, pp. 402–426). Hoboken, NJ: Wiley.
- Nguyen, L., Huang, L. N., Arganza, G. F., & Liao, Q. (2007). The influence of race and ethnicity on psychiatric diagnoses and clinical characteristics of children and adolescents in children's services. *Cultural Diversity and Ethnic Minority Psychology*, 13, 18–25.
- Noh, S., & Avison, W. R. (1992). Assessing psychopathology in Korean immigrants: Some preliminary results on the SCL-90. *Canadian Journal of Psychiatry*, 37(9), 640–645.
- Norcross, J. C. (1997). Emerging breakthroughs in psychotherapy integration: Three predictions and one fantasy. *Psychotherapy*, 34, 86–90.
- Norcross, J. C. (2006). Integrating self-help into psychotherapy: 16 practical suggestions. *Professional Psychology: Research & Practice*, 37, 683–693.
- Norcross, J. C. (Ed.) (2011). *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed.). New York, NY: Oxford University Press.
- Norcross, J. C., Campbell, L. F., Grohol, J. M., Santrock, J. W., Selagea, F., & Sommer, R. (2013). *Self-help that works: Resources to improve emotional health and strengthen relationships* (4th ed.). New York, NY: Oxford University Press.
- Norcross, J. C., Krebs, P. M., & Prochaska, J. O. (2011). Stages of change. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 279–300). New York, NY: Oxford University Press.
- Norcross, J., Santrock, J. W., Campbell, L. F., Smith, T. P., Sommer, R., & Zuckerman, E. L. (2012). *Authoritative guide to self-help resources in mental health* (4th ed.). New York, NY: Guilford Press.
- Norman, S., Tröster, A. I., Fields, J. A., & Brooks, R. (2002). Effects of depression and Parkinson's disease on cognitive functioning. *Journal of Neuropsychiatry and Clinical Neuroscience*, 14(1), 31–36.
- Novy, D. M., Nelson, D. V., Smith, K. G., Rogers, P. A., & Rowzee, R. D. (1995). Psychometric comparability of the English- and Spanish-language versions of the State-Trait Anxiety Inventory. *Hispanic Journal of Behavioral Sciences*, 17(2), 209–224.
- Nuevo, R., Dunn, G., Dowrick, C., Vázquez-Barquero, J. L., Casey, P., Dalgard, O. S., ... Ayuso-Mateos, J. L. (2009). Cross-cultural equivalence of the Beck Depression Inventory:

- A five-country analysis from the ODIN study. *Journal of Affective Disorders*, 114(1), 156–162.
- Oas, P. (1984). Validity of the Draw-A-Person and Bender Gestalt as measures of impulsivity with adolescents. *Journal of Consulting and Clinical Psychology*, 52, 1011–1019.
- O'Brien, W. H., Kaplar, M. E., & McGrath, J. J. (2004). Broadly based causal models of behavior disorders. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3, pp. 69–93). Hoboken, NJ: Wiley.
- O'Donohue, W. T., & Fisher, J. E. (2006). Clinician's handbook of evidence-based practice guidelines: The role of practice guidelines in systematic quality improvement. In J. E. Fisher & W. T. O'Donohue (Eds.), *Practitioner's guide to evidence-based psychotherapy* (pp. 1–23). New York, NY: Springer.
- Office of Science and Technology. (1967). *Privacy and behavioral research*. Washington, DC: U.S. Government Printing Office.
- Office of Strategic Services Staff. (1948). *Assessment of men*. New York, NY: Holt, Rinehart and Winston.
- Okun, A., Stein, R. K., Bauman, L. J., & Silver, E. J. (1996). Content validity of the Psychiatric Symptom Index, CES Depression Scale, and State-Trait Anxiety Inventory from the perspective of DSM-IV. *Psychological Reports*, 79, 1059–1069.
- Ollendick, T. H. (1978). *The Fear Survey Schedule for Children—Revised*. Unpublished manuscript, Indiana State University, Terre Haute, IN.
- Ollendick, T. H. (1983). Reliability and validity of the revised Fear Survey Schedule for Children (FSSC-R). *Behavior Research and Therapy*, 21, 685–692.
- Ollendick, T. H., Alvarez, H. K., & Greene, R. W. (2004). Behavioral assessment: History of underlying concepts and methods. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Behavioral assessment* (Vol. 3, pp. 19–34). Hoboken, NJ: Wiley.
- Ollendick, T. H., & King, N. J. (2004). Empirically supported treatments for children and adolescents: Advances toward evidence-based practice. In P. M. Barrett & T. H. Ollendick (Eds.), *Handbook of interventions that work with children and adolescents: Prevention and treatment* (pp. 3–25). Hoboken, NJ: Wiley.
- Olsen, L. R., Mortensen, E. L., & Bech, P. (2004). The SCL-90 and SCL-90R versions validated by item response models in a Danish community sample. *Acta Psychiatrica Scandinavica*, 110(3), 225–229.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 766–783.
- Osman, A., Downs, W., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory–III. *Journal of Psychopathology and Behavioral Assessment*, 19, 359–376.
- Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and validity of the Beck Depression Inventory–II with adolescent psychiatric inpatients. *Psychological Assessment*, 16, 120–132.
- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. *Archives de Psychologie*, 30, 206–356.
- Othmer, E. O., & Othmer, S. C. (1994). *The clinical interview using DSM-IV: Vol. 1. Fundamentals*. Washington, DC: American Psychiatric Press.
- Othmer, E., & Othmer, S. C. (2002). *The clinical interview using DSM-IV: Vol. 1. Fundamentals*. (2nd ed.) Washington, DC: American Psychiatric Press.
- Otto, R. K., & Heilburn, K. (2002). The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist*, 57, 5–10.
- Ownby, R. L. (1990). A study of the expository process model in mental health settings. *Journal of Clinical Psychology*, 46, 366–371.
- Ownby, R. L. (1997). *Psychological reports: A guide to report writing in professional psychology* (3rd ed.). Brandon, VT: Clinical Psychology.

- Paap, M. C. S., Meijer, R. R., Cohen-Kettenis, P. T., Richter-Appelt, H., de Cuypere, G., Kreukels, B. P. C., ... Haraldsen, I. R. (2012). Why the factorial structure of the SCL-90-R is unstable: Comparing patient groups with different levels of psychological distress using Mokken scale analysis. *Psychiatry Research*, 200(2-3), 819-826.
- Pace, T. M., Robbins, R. R., Choney, S. K., Hill, J. S., Lacey, K., & Blair, G. (2006). A cultural-contextual perspective on the validity of the MMPI-2 with American Indians. *Cultural Diversity and Ethnic Minority Psychology*, 12(2), 320-333.
- Pachet, A. (2007). Construct validity of the Repeatable Battery of Neuropsychological Status (RBANS) with acquired brain injury patient. *Clinical Neuropsychologist*, 21, 286-297.
- Page, H. A. (1957). Studies in fantasy-daydreaming frequency and Rorschach scoring categories. *Journal of Consulting Psychology*, 21, 111-114.
- Pantle, M. L., Ebner, D. L., & Hynan, L. S. (1994). The Rorschach and the assessment of impulsivity. *Journal of Clinical Psychology*, 50, 633-638.
- Panzarella, C., Alloy, L., & Whitehouse, W. G. (2006). Extended hopelessness theory of depression: On the mechanisms by which social support protects against depression. *Cognitive Therapy and Depression*, 30, 307-333.
- Parker, K. C. H. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment*, 47, 227-231.
- Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367-373.
- Parks, C. W. (1982). *A multi-dimensional view of the imagery construct: Issues of definition and assessment*. Unpublished manuscript.
- Parks, C. W., & Hollon, S. D. (1988). Cognitive assessment. In A. S. Bellak & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed., pp. 161-212). New York, NY: Pergamon Press.
- Parry, G., & Shapiro, D. A. (1986). Social support and life events in working class women: Stress buffering or independent effects? *Archives of General Psychiatry*, 43(3), 315-323.
- Parsons, M. W., & Hammeke, T. E. (2014). *Clinical neuropsychology: A pocket handbook for assessment* (3rd ed.). Washington, DC: American Psychological Association.
- Pascal, G. R., & Suttell, B. J. (1951). *The Bender Gestalt Test: Quantification and validity for adults*. New York, NY: Grune & Stratton.
- Patterson, C. H. (1989). Foundations for a systematic eclectic psychotherapy. *Psychotherapy*, 26, 427-435.
- Patton, D. E., Duff, K., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. (2003). Performance of cognitively normal African Americans on the RBANS in community-dwelling older adults. *Clinical Neuropsychologist*, 17, 515-530.
- Paul, G. L. (1967). Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Handbook of psychotherapy integration* (pp. 300-334). New York, NY: Basic Books.
- Paul, G. L., & Lentz, R. J. (1977). *Psychological treatment of chronic mental patients: Milieu versus social-learning programs*. New York, NY: Harvard University Press.
- Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences*, 39(2), 297-308.
- Pauls, F., Petermann, F., & Lepach, A. C. (2013). Gender differences in episodic memory and visual working memory including the effects of age. *Memory*, 21(7), 857-874.
- Pauly, H., Linkersdörfer, J., Lindberg, S., Woerner, W., Hasselhorn, M., & Lonnemann, J. (2011). Domain-specific rapid automatized naming deficits in children at risk for learning disabilities. *Journal of Neurolinguistics*, 24(5), 602-610.
- Pearson. (2009a). *Advanced clinical solutions for the WAIS-IV and WMS-IV*. San Antonio, TX: Author.
- Pearson. (2009b). *Wechsler Individual Achievement Test* (3rd ed.). Minneapolis, MN: Author.
- Pearson. (2009c). *Wechsler Memory Scale—Fourth Edition administration and scoring manual*. San Antonio, TX: Author.

- Pearson. (2009d). *Wechsler Memory Scale—Fourth Edition technical and interpretive manual*. San Antonio, TX: Author.
- Pekarik, G., & Wolff, C. B. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice*, 27, 202–208.
- Penley, J. S., Wiebe, J. S., & Nwosu, A. (2003). Psychometric properties of the Spanish Beck Depression Inventory–II in a medical sample. *Psychological Assessment*, 15, 569–577.
- Perez, R. G., Ascaso, E. E., Massons, J. M. D., & Chaparro, N. (1998). Characteristics of the subject and interview influencing the test-retest reliability of the Diagnostic Interview for Children and Adolescents—Revised. *Journal of Child Psychology and Psychiatry*, 39, 963–972.
- Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist*, 52, 1125–1133.
- Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry*, 149, 1645–1653.
- Perry, W., Sprock, J., Schaible, D., McDougall, A., Minassian, A., Jenkins, M., & Braff, D. (1995). Amphetamine use on Rorschach measures in normal subjects. *Journal of Personality Assessment*, 64, 456–465.
- Persons, J. B. (2008). *The case formulation approach to cognitive-behavior therapy*. New York, NY: Guilford Press.
- Persons, J. B., Mooney, K., & Padesky, C. A. (1995). Interrater reliability of cognitive behavioral case formulations. *Cognitive Therapy and Research*, 19, 21–34.
- Petrocelli, J. V. (2002). Processes and stages of change: Counseling with the transtheoretical model of change. *Journal of Counseling and Development*, 80, 22–30.
- Peveler, R. C., & Fairburn, C. G. (1990). Measurement of neurotic symptoms by self report questionnaire: Validity of the SCL–90–R. *Psychological Medicine*, 20, 873–879.
- Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured interview for DSM-IV personality: SIDP–IV*. Arlington, VA: American Psychiatric Association.
- Piasecki, T. M., Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment*, 19, 25–43.
- Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T., Jr., (2002). The applicability of the five-factor model in Sub-Saharan culture: The NEO–PI–R in Shona. In R. R. McCrae & J. Alik (Eds.), *The five-factor model of personality across cultures* (pp. 155–173). New York, NY: Kluwer Academic/Plenum.
- Piedmont, R. L., & Chae, J. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO–PI–R for Koreans. *Journal of Cross-Cultural Psychology*, 28(2), 131–155.
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78(3), 582–593.
- Piedmont, R. L., Sokolov, R. L., & Flemming, M. Z. (1989a). An examination of some diagnostic strategies involving the Wechsler intelligence scales. *Psychological Assessment*, 1, 181–185.
- Piedmont, R. L., Sokolov, R. L., & Flemming, M. Z. (1989b). On WAIS-R difference scores in a psychiatric population. *Psychological Assessment*, 1, 155–159.
- Piersma, H. L. (1986). The stability of the Millon Clinical Multiaxial Inventory for psychiatric patients. *Journal of Personality Assessment*, 50, 193–197.
- Piersma, H. L., Boes, J. L., & Reaume, W. M. (1994). Unidimensionality of the Brief Symptom Inventory (BSI) in adult and adolescent inpatients. *Journal of Personality Assessment*, 63, 338–344.
- Pilowski, I., Spence, N., Cobb, J., & Katsikitis, M. (1984). The Illness Behavior Questionnaire as an aid in clinical assessment. *General Hospital Psychiatry*, 6, 123–130.

- Pincus, H.A., Pechura, C., Keyser, D., Bachman, J., & Houtsinger, J. K. (2006). Depression in primary care: Learning lessons in national quality improvement. *Administration and Policy in Mental Health and Mental Health Services Research*, 33, 2–15.
- Piotrowski, C. (1996). Use of the Beck Depression Inventory in clinical practice. *Psychological Reports*, 6, 74–82.
- Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology*, 55, 787–796.
- Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical training programs. *Journal of Personality Assessment*, 61, 394–405.
- Piotrowski, Z. A. (1957). *Perceptanalysis*. New York, NY: Macmillan.
- Piotrowski, Z. A. (1960). The movement score. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology* (pp. 49–58). New York, NY: Wiley.
- Piotrowski, Z. A., & Schreiber, M. (1952). Rorschach perceptanalytic measurement of personality changes during and after intensive psychoanalytically oriented psychotherapy. In B. Bychowski & J. L. Despert (Eds.), *Specialized techniques in psychotherapy* (pp. 337–362). New York, NY: Basic Books.
- Ponsford, J. L. (1988). Neuropsychological assessment: The need for a more pragmatic approach. *Australian Psychologist*, 23, 349–360.
- Ponsford, J. L. (2000). Attention. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 355–400). New York, NY: Wiley.
- Ponterotto, J. G., & Grieger, I. (2007). Effectively communicating qualitative research. *Counseling Psychologist*, 35, 404–430.
- Pope, K. S. (1992). Responsibilities in providing psychological test feedback to clients. *Psychological Assessment*, 4, 268–271.
- Pope, K. S. (2007a). Informed consent in psychotherapy and counseling: Forms, standards, guidelines, and references. Retrieved from <http://kspope.com/consent/index.php>.
- Pope, K. S. (2007b). Responsibilities in providing psychological test feedback to clients. Retrieved from <http://kspope.com/assess/feedabs1.php>.
- Pope, K. S., Butcher, J. N., & Seelen, J. (2000). *The MMPI, MMPI-2 and MMPI-A in court: A practical guide for expert witnesses and attorneys* (2nd ed.). Washington, DC: American Psychological Association.
- Porter, E. H. (1950). *An introduction to therapeutic counseling*. Boston, MA: Houghton-Mifflin.
- Poythress, N., Nicholson, R., Otto, R. K., Edens, J. F., Bonnie, R. J., Monahan, J., & Hoge, S. K. (1999). *The MacArthur Competence Assessment Tool—Criminal Adjudication: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Presley, G., Smith, C., Hilsenroth, M., & Exner, J. (2001). Rorschach validity with African Americans. *Journal of Personality Assessment*, 77, 491–507.
- Preston, J., O'Neal, J. H., & Talaga, M. C. (2005). *Handbook of clinical psychopharmacology for therapists* (4th ed.). Oakland, CA: New Harbinger.
- Pretzer, J. L., Epstein, N., & Fleming, B. (1992). The Marital Attitude Survey: A measure of dysfunctional attitudes and expectancies. *Journal of Cognitive Psychotherapy*, 5, 131–148.
- Price, L., Tulskey, D., Millis, S., & Weiss, L. (2002). Redefining the factor structure of the Wechsler Memory Scale—III: Confirmatory factor analysis with cross validation. *Journal of Clinical and Experimental Neuropsychology*, 24, 574–585.
- Prifitera, A., Weiss, L. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 1–38). New York, NY: Academic Press.
- Prigatano, G. P. (1986). *Neuropsychological rehabilitation after brain injury*. Baltimore, MD: Johns Hopkins University Press.
- Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1992a). *Manual for the BNI Screen for Higher Cerebral Functions*. Phoenix, AZ: Barrow Neurological Institute.
- Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1992b). Validity studies of the BNI Screen for Higher Cerebral Functions. *Barrow Neurological Institute Quarterly*, 9, 2–9.

- Prince, J. P., & Heiser, L. J. (2000). *Essentials of career interest assessment*. New York, NY: Wiley.
- Prinz, U., Nutzinger, D. O., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2013). Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders. *BMC Psychiatry*, 13(104), 1–9.
- Prochaska, J. O. (2000). Change at differing stages. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for the 21st century* (pp. 109–127). New York, NY: Wiley.
- Prochaska, J. O., & DiClemente, C. C. (1984). *The transtheoretical approach: Crossing the traditional boundaries of therapy*. Homewood, IL: Dow Jones-Irwin.
- Prochaska, J. O., & DiClemente, C. C. (2005). The transtheoretical approach. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed.; pp. 147–171). New York, NY: Oxford University Press.
- Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, 47, 1102–1114.
- Prochaska, J. O., Norcross, J. C., & DiClemente, C. C. (1994). *Changing for good*. New York, NY: Morrow.
- Prochaska, J. O., Rossi, J. S., & Wilcox, N. S. (1991). Change processes and psychotherapy outcome in integrative case research. *Journal of Psychotherapy Integration*, 1, 103–120.
- Pruitt, J. A., Smith, M. C., Thelen, M. H., & Lubin, B. (1985). Attitudes of academic clinical psychologists toward projective techniques: 1968–1983. *Professional Psychology: Research and Practice*, 16, 781–788.
- Prunas, A., Sarno, I., Preti, E., Madeddu, F., & Perugini, M. (2012). Psychometric properties of the Italian version of the SCL-90-R: A study on a large community sample. *European Psychiatry*, 27(8), 591–597.
- Psychological Corporation. (1997). *WAIS-III/WMS-III technical manual*. San Antonio, TX: Author.
- Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Author.
- Psychological Corporation (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Author.
- Puig-Antich, J., & Chambers, W. (1978). *The schedule for affective disorders and schizophrenia for school aged children*. New York, NY: New York State Psychiatric Institute.
- Putzke, J. D., Williams, M. A., Daniel, F. J., & Boll, T. (1999). The utility of K-Correction to adjust for defensive response set on the MMPI. *Assessment*, 6, 61–70.
- Quevedo, K. M., & Butcher, J. N. (2005). The use of the MMPI and MMPI-2 in Cuba: A historical overview from 1950 to the present. *International Journal of Clinical and Health Psychology*, 5(2), 335–347.
- Quillan, J., Besing, S., & Dinning, D. (1977). Standardization of the Rathus Assertiveness Schedule. *Journal of Clinical Psychology*, 33, 418–422.
- Quilty, L. C., Zhang, K. A., & Bagby, R. M. (2010). The latent symptom structure for the Beck Depression Inventory-II in outpatients with major depression. *Psychological Assessment*, 22(3), 603–608.
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Psychology*, 20, 33–65.
- Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status manual*. San Antonio, TX: The Psychological Corporation.
- Randolph, C. (2012). *Repeatable Battery for the Assessment of Neuropsychological Status Update manual*. San Antonio, TX: Psychological Corporation.
- Rapaport, D., Gill, M., & Schafer, J. (1968). *Diagnostic psychological testing* (Vol. 1, rev. ed.). Chicago, IL: Year Book.
- Rapport, L. J., Webster, J. S., & Dutra, R. L. (1994). Digit Span performance and unilateral neglect. *Neuropsychologia*, 32, 517–525.

- Rathus, S. A. (1972). An experimental investigation of assertive training in a group setting. *Journal of Behavior Therapy and Experimental Psychiatry*, 3, 81–86.
- Rathus, S. A. (1973). A 30-item schedule for assessing assertive behavior. *Behavior Therapy*, 4, 398–406.
- Rathus, S. A., & Nevid, J. S. (1977). Concurrent validity of the 30-item assertiveness schedule with a psychiatric population. *Behavior Therapy*, 8, 393–397.
- Reese, J. B., Viglione, D. J., & Giromini, L. (2014). A comparison between Comprehensive System and an early version of the Rorschach Performance Assessment System administration with outpatient children and adolescents. *Journal of Personality Assessment*, 96(5), 515–522.
- Reich, J. H., & Noyes, R. (1987). A comparison of DSM-III personality disorders in acutely ill panic and depressed patients. *Journal of Anxiety Disorders*, 1, 123–131.
- Reich, W. (2000). Diagnostic Interview for Children and Adolescents (DICA). *Journal of the Academy of Child and Adolescent Psychiatry*, 39, 59–66.
- Reid, D. B., & Kelly, M. P. (1993). Wechsler Memory Scale–Revised in closed head injury. *Journal of Clinical Psychology*, 49, 245–254.
- Reinecke, M. A., Beebe, D. W., & Stein, M. A. (1999). The third factor of the WISC-III: It's (probably) not freedom from distractibility. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 322–328.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559.
- Reitan, R. M. (1955a). Certain differential effects of left and right cerebral lesions in human adults. *Journal of Comparative and Physiological Psychology*, 48, 474–477.
- Reitan, R. M. (1955b). Validity of the Rorschach Test as a measure of the psychological effects of brain damage. *Archives of Neurology and Psychiatry*, 73, 445–451.
- Reitan, R. M. (1974). Psychological effects of cerebral lesions in children of early school age. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and applications* (pp. 53–90). Washington, DC: Winston & Sons.
- Reitan, R. M., & Wolfson, D. (1985). *The Halsted-Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. Tucson, AZ: Tucson Neuropsychology Press.
- Reitan, R. M., & Wolfson, D. (1992). *Neuropsychological education of older children*. Tucson, AZ: Neuropsychology Press.
- Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press.
- Retzlaff, P. D. (1995). *Tactical psychotherapy of the personality disorders: An MCMI-III-based approach*. Boston, MA: Allyn & Bacon.
- Retzlaff, P. D., & Dunn, T. (2003). The Millon Clinical Multiaxial Inventory. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (pp. 192–226). New York, NY: Guilford Press.
- Retzlaff, P. D., Lorr, M., & Hyer, L. (1989). *An MCMI-II item-level component analysis: Personality and clinical factors*. Unpublished manuscript.
- Retzlaff, P. D., Sheehan, E. P., & Fiel, A. (1991). MCMI-II report style and bias: Profile and validity scales analysis. *Journal of Personality Assessment*, 56, 478–486.
- Rey, A. (1941). Psychological examination of traumatic encephalopathy. *Archives de Psychologie*, 28, 286–340.
- Rey, A. (1964). *The clinical exam in psychology*. Paris, France: Presses Universitaires de France.
- Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, 6, 144–150.
- Reynolds, C. R. (2007). *The Koppitz Developmental System for the Bender Gestalt Test* (2nd ed.). Austin, TX: PRO-ED.
- Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology*, 25, 323–324.

- Reynolds, C. R., & Hartlage, L. (1979). Comparison of WISC and WISC-R regression lines for academic prediction with black and white referred children. *Journal of Consulting and Clinical Psychology, 47*, 589–591.
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children* (3rd ed.). Minneapolis, MN: Pearson Assessments.
- Reynolds, S. K., & Clark, L. A. (2001). Predicting dimensions of personality disorder from domains and facets of the five factor model. *Journal of Personality, 69*(2), 199–222.
- Ridgeway, E. M., & Exner, J. E. (1980). *Rorschach correlates of achievement needs in medical students under an arousal state*. Rorschach Workshops (Study No. 274, unpublished).
- Ridley, C., & Kelly, S. (2007). Multicultural considerations in case formulation. In T. D. Eells (Ed.), *Handbook of psychotherapy case formulation* (2nd ed., pp. 33–64). New York, NY: Guilford Press.
- Riggs, D. S., & Foa, E. B. (1993). Obsessive compulsive disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (2nd ed., pp. 189–239). New York, NY: Guilford Press.
- Riskind, J. H., Beck, A. T., Berchick, R. J., Brown, G., & Steer, R. A. (1987). Reliability of DSM-III diagnoses for major depression and generalized anxiety disorder using the Structured Clinical Interview for DSM-III. *Archives of General Psychiatry, 44*, 817–820.
- Robin, R. W., Greene, R. L., Albaugh, B., Caldwell, A., & Goldman, D. (2003). Use of the MMPI-2 in American Indians: I. Comparability of the MMPI-2 between two tribes and with the MMPI-2 normative group. *Psychological Assessment, 15*, 351–359.
- Robins, L. N., Cottler, L. B., Bucholz, K. K., & Compton, W. (1996). *The Diagnostic Interview Schedule (Version IV)*. St. Louis, MO: Washington University School of Medicine.
- Robins, L. N., & Helzer, J. E. (1994). The half life of a structured interview: The NIMH Diagnostic Interview Schedule (DIS). *International Journal of Methods in Psychiatric Research, 4*, 95–102.
- Robins, L. N., Helzer, J. E., Cottler, L. B., & Goldring, E. (1989). *NIMH Diagnostic Interview Schedule (Version III–Revised)*. St. Louis, MO: Washington University School of Medicine.
- Robins, L. N., Helzer, J. E., Croughan, J. L., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule. *Archives of General Psychiatry, 38*, 381–389.
- Robinson, D. J. (2001). *Brain calipers: Descriptive psychopathology and the mental status examination* (2nd ed.). Port Huron, MI: Rapid Psychler Press.
- Robinson, J. D., & Baker, J. (2006). Psychological consultation and services in a general medical hospital. *Professional Psychology: Research and Practice, 37*, 264–267.
- Rock, R. C., Sellbom, M., Ben-Porath, Y. S., & Salekin, R. T. (2013). Concurrent and predictive validity of psychopathy in a batterers' intervention sample. *Law and Human Behavior, 37*(3), 145–154.
- Rodenhauser, P., & Fornal, R. E. (1991). How important is the mental status examination? *Psychiatric Hospital, 22*, 256–262.
- Rodgers, D. A. (1972). The MMPI: A review. In O. K. Buros (Ed.), *Seventh mental measurements yearbook* (Vol. 1, pp. 243–250). Highland Park, NJ: Gryphon Press.
- Rogers, C. R. (1961). *On becoming a person*. Boston, MA: Houghton Mifflin.
- Rogers, C. R. (1992). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting and Clinical Psychology, 60*(6), 827–832. (Original work published 1957).
- Rogers, R. (1984). *Rogers criminal responsibility assessment scales*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York, NY: Guilford Press.
- Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment, 60*, 215–226.
- Rogers, R., Flores, J., Ustad, K., & Sewell, K. W. (1995). Initial validation of the Personality Assessment Inventory–Spanish Version with clients from Mexican American communities. *Journal of Personality Assessment, 64*(2), 340–348.

- Rogers, R., Jackson, R. L., & Cashiel, M. (2004). The Schedule for Affective Disorders and Schizophrenia (SADS). In M. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 144–152). Hoboken, NJ: Wiley.
- Rogers, R., Salekin, R. T., & Sewell, K. W. (1999). Validation of the Millon Clinical Multiaxial Inventory for Axis II disorders: Does it meet the Daubert standard? *Law and Human Behavior*, 22, 425–443.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10, 160–177.
- Rogers, R., Sewell, K. W., Morey, L. C., & Ulstad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67(3), 629–640.
- Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of Competence to Stand Trial—Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R., Ulstad, K. L., & Salekin, R. T. (1998). Convergent validity of the Personality Assessment Inventory: A study of emergency referrals in a correctional setting. *Assessment*, 5(1), 3–12.
- Roid, G. H., Prifitera, A., & Ledbetter, M. (1988). Confirmatory analysis of the factor of the Wechsler Memory Scale–Revised. *Clinical Neuropsychologist*, 2, 116–120.
- Romanczyk, R. G., Kent, R. N., Diament, C., & O’Leary, K. D. (1973). Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis*, 6, 175–184.
- Rorschach, H. (1941). *Psychodiagnostik* (Hans Huber Verlag, Trans.). Bern, Switzerland: Bircher. (Original work published 1921)
- Rosen, B. M., Bahn, A. K., & Kramer, M. (1964). Demographic and diagnostic characteristics of psychiatric clinic patients in the U.S.A., 1961. *American Journal of Orthopsychiatry*, 34, 455–468.
- Rosenhan, D. L. (1973). On being sane in insane places. *Science*, 179, 250–257.
- Rosenthal, M. J. (1989). Towards selective and improved performance of the mental status examination. *Acta Psychiatrica Scandinavica*, 80, 207–215.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York, NY: Appleton-Century-Crofts.
- Rosenthal, R., & Fode, K. L. (1963). The effects of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183–189.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York, NY: Holt, Rinehart and Winston.
- Rosqvist, J., Sundsmo, A., MacLane, C., Cullen, K., Norling, D. C., Davies, M., & Maack, D. (2006). Analogue and virtual reality assessment. In M. Hersen (Ed.), *Clinician’s handbook of adult behavioral assessment* (pp. 43–63). New York, NY: Elsevier.
- Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York, NY: Academic Press.
- Rossi, G., Elklit, A., & Simonsen, E. (2010). Empirical evidence for a four factor framework of personality disorder organization: Multigroup confirmatory factor analysis of the Millon Clinical Multiaxial Inventory–III personality disorder scales across Belgian and Danish data samples. *Journal of Personality Disorders*, 24, 128–150.
- Rossini, E. D., & Kaspar, J. C. (1987). The validity of the Bender-Gestalt emotional indicators. *Journal of Personality Assessment*, 51, 254–261.
- Rouse, S. V., Butcher, J. N., & Miller, K. B. (1999). Assessment of substance abuse in psychotherapy clients: The effectiveness of the MMPI-2 substance abuse scales. *Psychological Assessment*, 11, 101–107.
- Rude, S. R. (1986). Relative benefits of assertion or cognitive self-control treatment for depression as a function of proficiency in each domain. *Journal of Consulting and Clinical Psychology*, 54, 390–394.

- Ruegg, R. G., Ekstrom, D. E., Evans, D. L., & Golden, R. N. (1990). Introduction of a standardized report form improves the quality of mental status examination reports by psychiatric residents. *Academic Psychiatry, 14*, 157–163.
- Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology, 68*, 145–154.
- Rushton, J. P. (1994). The equalitarian dogma revisited. *Intelligence, 19*, 263–280.
- Rushton, J. P., & Irwing, P. (2009). A general factor of personality in 16 sets of the big five, the Guilford-Zimmerman Temperament Survey, the California Psychological Inventory, and the Temperament and Character Inventory. *Personality and Individual Differences, 47*(6), 558–564.
- Russ, S. W. (2001). Tackling ethical dilemmas in personality assessment. *Journal of Personality Assessment, 77*, 255–258.
- Russell, E. W. (1975). A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology, 43*, 800–809.
- Russell, E. W. (1988). Renorming Russell's version of the Wechsler Memory Scale. *Journal of Clinical and Experimental Neuropsychology, 10*, 235–249.
- Russell, E. W. (2000). The cognitive-metric, fixed battery approach to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 449–483). Hillsdale, NJ: Erlbaum.
- Ryan, C. M., Morrow, L., Parklinson, D., & Branet, E. (1987). Low level lead exposure and neuropsychological functioning in blue collar males. *International Journal of Neuroscience, 36*, 29–39.
- Ryan, J. J., Glass, L. A., & Brown, C. N. (2007). Administration time estimates for Wechsler Intelligence Scale for Children–IV subtests, composites, and short forms. *Journal of Clinical Psychology, 63*, 309–318.
- Ryan, J. J., Lopez, S. J., & Werth, T. R. (1998). Administration time estimates for WAIS-III subtests, scales, and short forms in a clinical sample. *Journal of Psychoeducational Assessment, 16*, 315–323.
- Ryan, J. J., Lopez, S. J., & Werth, T. R. (1999). Development and preliminary validation of a Satz-Mogel short form of the WAIS-III in a sample of persons with substance abuse disorders. *International Journal of Neurosciences, 98*, 131–140.
- Ryan, J. J., Paolo, A. M., & Brungardt, T. M. (1990). Standardization of the Wechsler Adult Intelligence Scale–Revised for persons 75 years and older. *Psychological Assessment, 2*, 404–411.
- Ryan, J. J., & Paul, C. A. (1999). Who is president of the United States? *Perceptual and Motor Skills, 89*, 595–596.
- Ryan, J. J., Paul, C. A., & Arb, J. D. (1999). Intrasubtest scatter on the WAIS-III Information subtest and psychometrically defined retrieval deficits. *Perceptual and Motor Skills, 89*, 1052–1058.
- Ryan, J. J., & Schnakenberg-Ott, S. D. (2003). Scoring reliability on the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III). *Assessment, 10*(2), 151–159.
- Ryan, J. J., & Ward, L. C. (1999). Validity, reliability, and standard errors of measurement for two seven subtest short forms of the Wechsler Adult Intelligence Scale–III. *Psychological Assessment, 11*, 207–211.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising evidence for validity and fairness. *American Psychologist, 63*, 215–227.
- Sadock, B. J., & Sadock, V. A. (2010). *Kaplan and Sadock's pocket handbook of clinical psychiatry* (5th ed.). Philadelphia, PA: Lippincott Williams and Wilkins.
- Sales, J., & Miller, P. (1994). *Psychology in litigation and legislation*. Washington, DC: American Psychological Association.
- Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64*, 80–97.

- Sammons, M. T., & Schmidt, N. B. (2001). *Combined treatment for mental disorders: A guide to psychological and pharmacological interventions*. Washington, DC: American Psychological Association.
- Sanderson, C., & Clarkin, J. F. (2002). Further use of the NEO-PI-R personality dimensions in differential treatment planning. In P. T. Costa Jr., & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (2nd ed., pp. 351–375). Washington, DC: American Psychological Association.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology*, 47, 919–927.
- Sandy, L. R. (1986). The descriptive-collaborative approach to psychological report writing. *Psychology in the Schools*, 23, 395–400.
- Saran, M., Phansalkar, S., & Kablinger, A. S. (2007). Biological markers and the future of early diagnosis and treatment of schizophrenia. *Psychiatric Times*, 24, 1–3.
- Sarason, I. G., Levine, H. M., Basham, R. B., & Sarason, B. R. (1983). Assessing social support: The Social Support Questionnaire. *Journal of Personality and Social Psychology*, 44, 127–139.
- Sattler, J. M. (1973a). Examiners' scoring style, accuracy, ability, and culturally disadvantaged children. In L. Mann & D. Sabatino (Eds.), *The first review of special education* (Vol. 2). Philadelphia, PA: J. S. E. Press.
- Sattler, J. M. (1973b). Racial experimenter effects. In K. S. Miller & R. M. Dreger (Eds.), *Comparative studies of blacks and whites in the United States* (pp. 8–32). New York, NY: Seminar Press.
- Sattler, J. M. (1985). Review of the Hutt adaptation of the Bender-Gestalt Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 1, pp. 184–185). Highland Park, NJ: Gryphon Press.
- Sattler, J. M. (2001). *Assessment of children: Cognitive functions* (4th ed.). San Diego, CA: Author.
- Sattler, J. M. (2008). *Assessment of children: Cognitive functions* (5th ed.). San Diego, CA: Author.
- Sattler, J. M. (2014). *Foundations of behavioral, social, and clinical assessment of children* (6th ed.). San Diego, CA: Author.
- Sattler, J. M., & Gwynne, J. (1982). White examiners generally do not impede the intelligence test performance of black children: To debunk a myth. *Journal of Consulting and Clinical Psychology*, 50, 196–208.
- Sattler, J. M., Hillix, W. A., & Neher, L. A. (1970). Halo effect in examiner scoring of intelligence test responses. *Journal of Consulting and Clinical Psychology*, 34, 172–176.
- Sattler, J. M., & Hoge, R. D. (2006). *Assessment of children: Behavioral, social and clinical foundations* (5th ed.). San Diego, CA: Jerome M. Sattler.
- Sattler, J. M., & Winget, B. M. (1970). Intelligence testing procedures as affected by expectancy and I.Q. *Journal of Clinical Psychology*, 26, 446–448.
- Satz, P., & Mogel, S. (1962). An abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology*, 18, 77–79.
- Sayers, S. L., & Tomcho, T. J. (2006). Behavioral interviewing. In M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 63–84). New York, NY: Elsevier.
- Sbordone, R. J. (2000a). The assessment interview in clinical neuropsychology. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 94–128). New York, NY: Wiley.
- Sbordone, R. J. (2000b). The executive functions of the brain. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 437–456). New York, NY: Wiley.
- Sbordone, R. J., & Guilmette, T. J. (1999). Ecological validity: Prediction of everyday and vocational functioning from neuropsychological test data. In J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 223–250). New York, NY: Swets.

- Sbordone, R. J., & Long, C. J. (Eds.). (1996). *Ecological validity of neuropsychological testing*. Odessa, FL: Psychological Assessment Resources.
- Scandell, D. J. (2000). Development and initial validation of validity scales for the NEO-Five Factor Inventory. *Personality and Individual Differences*, 29(6), 1153–1162.
- Scepansky, J. A., & Bjornsen, C. A. (2003). Educational, orientation, NEO PI-R personality traits, and plans for graduate school. *College Student Journal*, 37(4), 574–581.
- Schalock, R. L., Luckasson, R. A., Shogren, K. A., Borthwick-Duffy, S., Bradley, V., Buntinx, W. H. E., ... & Yaeger, M. H. (2007). The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities*, 45, 116–124.
- Schinka, J. A. (1983). *Neuropsychological Status Examination*. Odessa, FL: Psychological Assessment Resources.
- Schinka, J. A. (1995). Personality Assessment Inventory scale characteristics and factor structure in the assessment of alcohol dependency. *Journal of Personality Assessment*, 64(1), 101–111.
- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment*, 68(1), 127–138.
- Schinka, J. A., LaLone, L., & Greene, R. (1998). Effects of psychopathology and demographic characteristics on MMPI-2 scale scores. *Journal of Personality Assessment*, 70, 197–211.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the work place. *Journal of Personality and Social Psychology*, 86, 162–173.
- Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology*, 43, 627–670.
- Schmidt, H. O., & Fonda, C. P. (1954). Rorschach scores in the manic state. *Journal of Psychology*, 38, 427–437.
- Schmidt, M., Trueblood, W., Merwin, M., & Durham, R. L. (1994). How much do “attention” tests tell us? *Archives of Clinical Neuropsychology*, 9, 383–394.
- Schmitz, N., & Hartkamp, N. (2000). Assessing clinically significant change: Application to the SCL-90-R. *Psychological Reports*, 86, 263–274.
- Schmitz, N., Hartkamp, N., Kiuse, J., Franke, G. H., Reister, G., & Tress, W. (2000). The Symptom Check-List-90-R (SCL-90-R): A German validation study. *Quality of Life Research*, 9(2), 185–193.
- Schmitz, N., Kruse, J., Heckrath, L., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: The General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology*, 34, 360–366.
- Schoenberg, M. R., Dorr, D., & Morgan, C. D. (2003). The ability of the Millon Clinical Multiaxial Inventory-Third Edition to detect malingering. *Psychological Assessment*, 15, 198–204.
- Schoenberg, M. R., Lange, R. T., Brickell, T. A., & Saklofske, D. H. (2007). Estimating premorbid general cognitive functioning for children and adolescents using the American Wechsler Intelligence Scale for Children-Fourth Edition: Demographics and current performance approaches. *Journal of Child Neurology*, 22, 379–388.
- Schoevers, R. A., Beekman, A. T. F., Deeg, D. J. H., Jonker, C., & Tilburg, W. V. (2003). Comorbidity and risk-patterns of depression, generalised anxiety disorder and mixed anxiety-depression in later life: Results from the AMSTEL study. *International Journal of Geriatric Psychiatry*, 18(11), 994–1001.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson Psychoeducational Battery-IV*. Itasca, IL: Riverside.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, 73, 305–316.

- Schuerger, J. M., & Witt, A. C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology, 45*, 294–302.
- Schulman, I. (1953). *The relation between perception of movement on the Rorschach test and levels of conceptualization: An experimental study and theoretical analysis of thought processes involved in the perception of movement on the Rorschach test* (Unpublished doctoral dissertation). New York University, New York.
- Schultz, C. B., & Sherman, R. H. (1976). Social class, development, and differences in reinforcer effectiveness. *Review of Educational Research, 46*, 25–59.
- Schwartz, G. E. (1982). Testing the biopsychosocial model: The ultimate challenge facing behavioral medicine? *Journal of Consulting and Clinical Psychology, 50*, 1040–1053.
- Schwartz, L., & Levitt, E. E. (1960). Short forms of the WISC for children in the educable, non-institutionalized mentally retarded. *Journal of Educational Psychology, 51*, 187–190.
- Schwartz, S., & Wiedel, T. C. (1981). Incremental validity of the MMPI in neurological decision-making. *Journal of Personality Assessment, 45*, 424–426.
- Segal, D. L., Coolidge, F. L., Cahill, B. S., & O'Riley, A. A. (2008). Psychometric properties of the Beck Depression Inventory–II (BDI–II) among community-dwelling older adults. *Behavioral Modification, 32*(1), 3–20.
- Seligman, L., & Reichenberg, L. W. (2014). *Selecting effective treatments: A comprehensive systematic guide to treating mental disorders* (4th ed.). Hoboken, NJ: Wiley.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*, 965–974.
- Seligman, M. E. P., Abramson, L. Y., Semmel, A., & von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology, 88*, 242–247.
- Seligman, M. E. P., Steen, T., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist, 60*, 410–421.
- Sellbom, M., Ben-Porath, Y., Baum, L. J., Erez, E., & Gregory, C. (2008). Predictive validity of the MMPI-2 restructured clinical (RC) scales in a batterers' intervention program. *Journal of Personality Assessment, 90*(2), 129–135.
- Sellbom, M., Ben-Porath, Y., & Graham, J. (2006). Correlates of the MMPI-2 Restructured Clinical (RC) scales in a college counseling setting. *Journal of Personality Assessment, 86*, 89–99.
- Sellbom, M., Graham, J. R., & Schenk, P. W. (2006). Incremental validity of the MMPI-2 restructured clinical (RC) scales in a private practice sample. *Journal of Personality Assessment, 86*(2), 196–205.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2–RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment, 22*(1), 22–31.
- Sellbom, M., Wygant, D. B., & Bagby, M. (2012). Utility of the MMPI-2–RF in detecting non-credible somatic complaints. *Psychiatry Research, 197*(3), 295–301.
- Sewitch, T., & Kirsch, I. (1984). The cognitive content of anxiety: Naturalistic evidence for the predominance of threat-related thoughts. *Cognitive Therapy and Research, 8*, 49–58.
- Shaevel, B., & Archer, R. P. (1996). Effects of MMPI-2 and MMPI-A norms on t-score elevations for 18-year-olds. *Journal of Personality Assessment, 67*(1), 72–78.
- Shaffer, D., Fisher, P. W., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children (Version IV; NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnosis. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 28–38.
- Shalit, B. (1965). Effects of environmental stimulation on the M, FM, and m response to the Rorschach. *Journal of Projective Techniques and Personality Assessment, 29*, 228–231.
- Shanahan, S. F., Anderson, S. J., & Mkhize, N. J. (2001). Assessing psychological distress in Zulu-speakers: Preliminary findings from an adaptation of the SCL-90–R. *South African Journal of Psychology, 31*(4), 1–12.
- Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral

- and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 522–534.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139–156.
- Shatin, L. (1952). Psychoneurosis and psychosomatic reactions: A Rorschach contrast. *Journal of Consulting Psychology*, 16, 220–223.
- Shaver, P. R., & Brennan, K. A. (1992). Attachment styles and the “Big Five” personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin*, 18(5), 536–545.
- Shear, M. K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., & Hanekamp, M. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry*, 157, 581–587.
- Shedler, J., & Westen, D. (2004). Refining personality disorder diagnosis: Integrating science and practice. *American Journal of Psychiatry*, 161(8), 1350–1365.
- Sheikh, A. A. (2003). *Healing images: The role of imagination in health*. Amityville, NY: Baywood.
- Shen, L., & Dillard, J. P. (2007). Reactance proneness assessment. In R. A. Reynolds, R. Woods, & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 323–329). Hershey, PA: Idea Group Reference.
- Sherman, E. M. S., Strauss, E., Slick, D. J., & Spellacy, F. (2000). Effect of depression on neuropsychological functioning in head injury: Measurable but minimal. *Brain Injury*, 14, 621–632.
- Sheslow, D., & Adams, W. (2003). *WRAML-II manual*. Wilmington, DE: Jastak Associates.
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298–311.
- Shkalim, E. (2015). Psychometric evaluation of the MMPI-2/MMPI-2–RF restructured clinical scales in an Israeli sample. *Assessment*, 22(5), 607–618.
- Shorkey, C. L., Reyes, E., & Whiteman, V. L. (1977). Development of the rational behavior inventory: Initial validity and reliability. *Educational and Psychological Measurement*, 37, 527–534.
- Shrauger, J. S., & Osberg, T. M. (1981). The relative accuracy of self-predictions and judgments of others in psychological assessment. *Psychological Bulletin*, 90, 322–351.
- Sieber, K. O., & Meyers, L. S. (1992). Validation of the MMPI-2 social introversion subscales. *Psychological Assessment*, 4, 185–189.
- Siefert, C. J., Sinclair, S. J., Kehl-Fie, K. A., & Blais, M. A. (2009). An item-level psychometric analysis of the Personality Assessment Inventory Clinical scales in a psychiatric inpatient unit. *Assessment*, 16(4), 373–383.
- Siegel, A. W., Schechter, M. D., & Diamond, S. P. (1996). Neurobehavioral assessment format. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 429–504). Delray Beach, FL: St. Lucie Press.
- Siegler, I. C., Zonderman, A. B., Barefoot, J. C., Williams, R. B., Jr., Costa, P. T., Jr., & McCrae, R. R. (1990). Predicting personality in adulthood from college MMPI scores: Implications for follow-up studies in psychosomatic medicine. *Psychosomatic Medicine*, 52(6), 644–652.
- Sigelman, B. E., Spanhel, C., & Schoenrock, C. (1981, April). When in doubt say yes: Acquiescence in interviews with mentally retarded persons. *Mental Retardation*, 53–58.
- Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment*, 2, 3–11.
- Simon, R. I., & Hales, R. E. (Eds.). (2012). *The American Psychiatric Publishing textbook of suicide assessment and management*. Arlington, VA: American Psychiatric Association.
- Simon, W. E. (1969). Expectancy effects in the scoring of vocabulary items: A study of scorer bias. *Journal of Educational Measurement*, 6, 159–164.

- Simons, A. D., & Thase, M. E. (1992). Biological markers, treatment outcome, and 1 year follow-up in endogenous depression: Electroencephalographic sleep studies and response to cognitive therapy. *Journal of Consulting and Clinical Psychology*, 60, 392–401.
- Simons, R., Goddard, R., & Patton, W. (2002). Hand-scoring error rates in psychological testing. *Assessment*, 9(3), 292–300.
- Singer, J. A. (2005). *Personality and psychotherapy: Treating the whole person*. New York, NY: Guilford Press.
- Singer, J. L. (1960). The experience type: Some behavioral correlates and theoretical implications. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York, NY: Wiley.
- Singer, R. M. (1990). *Neurotoxicity guidebook*. New York, NY: Van Nostrand-Reinhold.
- Slate, R. J., & Hunnicutt, L. C. (1988). Examiner errors on the Wechsler scales. *Journal of Psychoeducational Assessment*, 6, 280–288.
- Slate, R. J., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale–Revised: An empirical evaluation of practice administrations. *Professional Psychology*, 22, 375–379.
- Smith, C. P., & Graham, J. R. (1981). Behavioral correlates for the MMPI standard F scale and the modified F scale for black and white psychiatric patients. *Journal of Consulting and Clinical Psychology*, 49, 455–459.
- Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the Draw-A-Person Test. *Professional Psychology: Research and Practice*, 26, 298–303.
- Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, 17, 396–408.
- Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300–308.
- Smith, M. L., Glass, G. V., & Miller, T. L. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
- Smith, S. R., Baity, M. R., Knowles, E. S., & Hilsenroth, M. J. (2002). Assessment of disordered thinking in children and adolescents: The Rorschach Perceptual-Thinking Index. *Journal of Personality Assessment*, 76, 333–351.
- Smith, S. R., Hilsenroth, M. J., Castlebury, F. D., & Durham, T. W. (1999). The clinical utility of the MMPI-2 Antisocial Practices Content Scale. *Journal of Personality Disorders*, 13, 385–393.
- Smith, T. B., Domenech Rodriguez, M., & Bernal, G. (2011). Culture. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 316–335). New York, NY: Oxford University Press.
- Snyder, C. R., Ritschel, L. A., Rand, K. L., & Berg, C. J. (2006). Balancing psychological assessments: Including strengths and hope in client reports. *Journal of Clinical Psychology*, 62(1), 33–46.
- Snyder, P. J., Nussbaum, P. D., & Robins, D. L. (Eds.). (2006). *Clinical neuropsychology: A pocket handbook for assessment* (2nd ed.). Washington, DC: American Psychological Association.
- Snyder, W. V. (1945). An investigation of the nature of nondirective psychotherapy. *Journal of General Psychology*, 33, 139–223.
- Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137–144.
- Sobel, D. S. (2000). Mind matters, money matters: The cost effectiveness of mind-body medicine. *Journal of the American Medical Association*, 284, 1705.
- Sommers-Flanagan, J., & Sommers-Flanagan, R. (2013). *Clinical interviewing* (5th ed.). Hoboken, NJ: Wiley.
- Sotsky, S. M., Glass, D. R., Shea, M. T., Pilkonis, P. A., Collins, F., Elkin, I., ... Oliveri, M. E. (2006). Patient predictors of response to psychotherapy and pharmacotherapy: Findings in the NIMH Treatment of Depression Collaborative Research Program. *Focus*, 4(2), 278–290.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales*. Minneapolis, MN: Pearson Clinical.

- Spielberger, C. D. (1973). *Manual for the State-Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D., Moscoso, M. S., & Brunner, T. M. (2004). Cross-cultural assessment of emotional states and personality traits. In R. K. Hambleton, C. D. Spielberger, & P. F. Merenda (Eds.), *Adapting educational and psychological tests for cross cultural assessment* (pp. 343–368). Mahwah, NJ: Erlbaum.
- Spielberger, C. D., & Piotrowski, C. (1990). Clinicians' attitudes toward computer-based testing. *Clinical Psychologist*, 43, 60–63.
- Spielberger, C. D., & Reheiser, E. C. (2004). Measuring anxiety, anger, depression, and curiosity as emotional states and personality traits with the STAI, STAXI, and STPI. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 70–86). Hoboken, NJ: Wiley.
- Spielberger, C. D., & Sydeman, S. J. (1994). State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment*. (pp. 292–321). Hillsdale, NJ: Erlbaum.
- Spielberger, C. D., Sydeman, S. J., Owen, A. E., & Marsh, B. J. (1999). Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed., pp. 993–1021). Mahwah, NJ: Erlbaum.
- Spiker, D. G., & Ehler, J. G. (1984). Structured psychiatric interviews for adult. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (pp. 291–304). New York, NY: Pergamon Press.
- Spitzer, R. L., Endicott, J., & Cohen, J. (1974). Constraints on the validity of computer diagnosis. *Archives of General Psychiatry*, 31, 197–203.
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, 35, 773–782.
- Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1987). *Structured clinical interview for DSM-III-R (SCID)*. New York, NY: State Psychiatric Institute.
- Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured clinical interview for DSM-III-R personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press.
- Springer, S. P., & Deutsch, G. (1998). *Left brain right brain: Perspectives from cognitive neuroscience* (5th ed.). New York, NY: Freeman.
- Stanley, M. A., Beck, J. G., & Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behaviour Research and Therapy*, 34, 827–838.
- Stanton, H. C., & Reynolds, C. R. (1998). *Configural frequency analysis as a method of determining Wechsler profile types*. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.
- Starcevic, V., Bogojevic, G., & Marinkovic, J. (2000). The SCL-90-R as a screening instrument for severe personality disturbance among outpatients with mood and anxiety disorders. *Journal of Personality Disorders*, 14, 199–207.
- Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1), 138–161.
- Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, 55, 117–118.
- Steer, R. A., Beck, A. T., & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 121–142). Geneva, Switzerland: World Health Organization.
- Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent depressed outpatients. *Journal of Psychopathology and Behavioral Assessment*, 20, 127–137.

- Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory–II with depressed geriatric inpatients. *Behaviour Research and Therapy*, 38, 311–318.
- Stein, L. A. R., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Using the MMPI-2 to detect substance abuse in an outpatient mental health setting. *Psychological Assessment*, 11, 94–100.
- Stein, L. A. R., Graham, J. R., & Williams, C. L. (1995). Detecting fake-bad MMPI-A profiles. *Journal of Personality Assessment*, 65, 415–427.
- Steinberg, M. (1993). *Interviewer's guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders (SCID-D)*. Washington, DC: American Psychiatric Press.
- Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the structured clinical interview for DSM-III-R and clinical diagnosis. *Journal of Nervous and Mental Disease*, 183, 365–369.
- Steketee, G. (1994). Behavioral assessment and treatment planning with obsessive compulsive disorder: A review emphasizing clinical application. *Behavior Therapy*, 25, 613–633.
- Stern, R. A., & White, T. (2003). *Manual for the Neuropsychological Assessment Battery*. Lutz, FL: Psychological Assessment Resources.
- Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. New York, NY: Cambridge University Press.
- Sternberg, R. J., & Grigorenko, E. (2001). *Dynamic testing: The nature and measurement of learning potential*. New York, NY: Cambridge University Press.
- Sternberg, R. J., & Kaufman, J. C. (1998). Innovation and intelligence testing: The curious case of the dog that didn't bark. *European Journal of Psychological Assessment*, 12, 175–182.
- Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. Testing common sense. *American Psychologist*, (1995), 50(11), 912–927.
- Stompe, T., Ortwein-Swoboda, G., Strobl, R., & Freidman, A. (2000). The age of onset of schizophrenia and the theory of anticipation. *Psychiatry Research*, 93, 125–134.
- Storandt, M., Botwinick, J., & Danzinger, W. L. (1986). Longitudinal changes: Patients with mild SDAT and matched health controls. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults*. Washington, DC: American Psychological Association.
- Stout, C. E., & Cook, L. P. (1999). New areas for psychological assessment in general health care settings: What to do today to prepare for tomorrow. *Journal of Clinical Psychology*, 55, 797–812.
- Strack, S. (1999). Millon's normal personality styles and dimensions. *Journal of Personality Assessment*, 72, 426–436.
- Strack, S. (2008). *Essentials of Millon inventories assessment* (3rd ed.). Hoboken, NJ: Wiley.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.
- Strauss, M. E., & Pasupathi, M. (1994). Primary caregivers' descriptions of Alzheimer patients' personality traits: Temporal stability and sensitivity to change. *Alzheimer Disease & Associated Disorders*, 8(3), 166–176.
- Stredny, R. V., & Ball, J. D. (2005). The utility of the Rorschach Coping Deficit Index as a measure of depression and social skills deficits in children and adolescents. *Assessment*, 12, 295–302.
- Streiner, D. L., & Miller, H. R. (1990). Maximum likelihood estimates of the accuracy of four diagnostic techniques. *Educational and Psychological Measurement*, 50, 653–662.
- Stringer, A. Y., & Nadolne, M. J. (2000). Neuropsychological assessment: Contexts for contemporary clinical practice. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 26–47). New York, NY: Wiley.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Strupp, H. H. (1958). The psychotherapist's contribution to the treatment process. *Behavioral Science*, 3, 34–67.

- Stukenberg, K. W., Brady, C., & Klinetob, N. (2000). Psychiatric inpatients and the MMPI-2: Providing benchmarks. *Journal of Clinical Psychology*, 747–756.
- Stukenberg, K. W., Dura, J. R., & Kiecolt-Glaser, J. K. (1990). Depression screening scale validation in an elderly, community-dwelling population. *Psychological Assessment*, 2, 134–138.
- Subica, A. M., Fowler, J. C., Elhai, J. D., Frueh, B. C., Sharp, C., Kelly, E. L., & Allen, J. G. (2014). Factor structure and diagnostic validity of the Beck Depression Inventory–II with adult clinical inpatients: Comparison to a gold-standard diagnostic interview. *Psychological Assessment*, 26(4), 1106–1115.
- Sue, D. W., Bingham, R. P., Porché-Burke, L., & Vásquez, M. (1999). The diversification of psychology: A multicultural revolution. *American Psychologist*, 54(12), 1061–1069.
- Suen, H. K., & Rzasa, S. R. (2004). Psychometric foundations of behavioral assessment. In M. Hersen (Ed.), *Comprehensive handbook of psychological assessment (Vol. 3): Behavioral assessment* (pp. 37–56). Hoboken, NJ: Wiley.
- Summerfeldt, L. J., Kloosterman, P. H., & Antony, M. M. (2011). Structured and semistructured diagnostic interviews. In M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 95–140). New York, NY: Guilford Press.
- Sundberg, N. D. (1955). The acceptability of “fake” versus “bona fide” personality test interpretations. *Journal of Abnormal and Social Psychology*, 50, 145–147.
- Suzuki, L. A., & Ponterotto, J. G. (Eds.). (2007). *Handbook of multicultural assessment: Clinical, psychological, and educational applications*. San Francisco, CA: Jossey-Bass.
- Swartz, M. S., Blazer, D. G., George, L. K., Winfield, I., Zakris, J., & Dye, E. (1989). Identification of borderline personality with the NIMH Diagnostic Interview Schedule. *American Journal of Psychiatry*, 146, 200–205.
- Sweet, J. J., Moberg, P. J., & Suchy, Y. (2000). Ten-year follow-up survey of clinical neuropsychologists. Part I: Practices and beliefs. *Clinical Neuropsychologist*, 14, 18–37.
- Swift, J. K., Callahan, J. L., & Vollmer, B. M. (2011). Preferences. *Journal of Clinical Psychology*, 67(2), 155–165.
- Szasz, T. (1987). Justifying coercion through religion and psychiatry. *Journal of Humanistic Psychology*, 27, 158–174.
- Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, 52, 1–23.
- Tallent, N. (1988). *Psychological report writing* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Tallent, N. (1992). *The practice of psychological assessment*. Englewood Cliffs, NJ: Prentice-Hall.
- Tallent, N. (1993). *Psychological report writing* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Talley, P. F., Strupp, H. S., & Morey, L. C. (1990). Matchmaking in psychotherapy: Patient-therapist dimensions and their impact on outcome. *Journal of Consulting and Clinical Psychology*, 58, 182–188.
- Tarescavage, A. M., Luna-Jones, L., & Ben-Porath, Y. S. (2014). Minnesota Multiphasic Personality Inventory–2–Restructured Form (MMPI-2–RF) predictors of violating probation after felonious crimes. *Psychological Assessment*, 26(4), 1375–1380.
- Tarrier, N., & Johnson, J. (Eds.). (2006). *Case formulation in cognitive behavior therapy: The treatment of challenging and complex cases*. New York, NY: Routledge.
- Tasca, G. A., Wood, J., Demidenko, N., & Bissada, H. (2002). Using the PAI with an eating disordered population: Scale characteristics, factor structure, and differences among diagnostic groups. *Journal of Personality Assessment*, 79(2), 337–356.
- Tellegen, A., & Ben-Porath, Y. S. (1993). Code-type comparability of the MMPI and MMPI-2: Analysis of recent findings and criticisms. *Journal of Personality Assessment*, 61, 489–500.
- Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2–RF (Minnesota Multiphasic Personality Inventory–2 Restructured Form): Technical manual*. Minneapolis: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J., & Kaemmer, B. (2003). *The MMPI-2 restructured clinical (RC) scales*. Minneapolis: University of Minnesota Press.

- Teng, E. L., Chui, H. C., & Saperia, D. (1990). Senile dementia: Performance on a neuropsychological test battery. *Recent Advances in Cardiovascular Disease*, 11, 27–34.
- Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York, NY: McGraw-Hill.
- Terracciano, A., Costa, P. T., Jr., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin*, 32(8), 999–1009.
- Terrell, F., Taylor, J., & Terrell, S. L. (1978). Effects of types of social reinforcement on the intelligence test performance of lower-class black children. *Journal of Consulting and Clinical Psychology*, 46, 1538–1539.
- Thelen, M. H., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the Bulimia Test: The BULIT-R. *Psychological Assessment*, 3, 119–124.
- Tibon Czopp, S., Rothschild-Yakar, L., & Appel, L. (2012). Rorschach Comprehensive System (CS) reference data for Israeli adolescents. *Journal of Personality Assessment*, 94(3), 276–286.
- Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2. *Psychological Assessment*, 6, 212–217.
- Tolman, A. O., & Rotzien, A. L. (2007). Conducting risk evaluations for future violence: Ethical practice is possible. *Professional Psychology: Research and Practice*, 38, 71–79.
- Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-mental state examination (MMSE) and the Modified MMSE (3MS): Psychometric comparison and normative data. *Psychological Assessment*, 8, 48–59.
- Tomioka, M., Shimura, M., Hidaka, M., & Kubo, C. (2008). The reliability and validity of a Japanese version of Symptom Checklist 90 Revised. *Biopsychosocial Medicine*, 2, 19.
- Tranel, D. (1994). The release of psychological data to nonexperts: Ethical and legal considerations. *Professional Psychology: Research and Practice*, 25, 33–38.
- Trimble, J. E., & Mohatt, G. V. (2006). The virtuous and responsible researcher in another culture. In J. E. Tribble & C. B. Fisher (Eds.), *The handbook of ethical research with ethnocultural populations and communities* (pp. 214–334). Thousand Oaks, CA: Sage.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy*. New York, NY: Aldine.
- Trumbetta, S. L., Bolinsky, P. K., & Gottesman, I. I. (2013). The MMPI-ARC and MMPI-AF scales: Psychometric properties and the MMPI-2-RF's restructured clinical (RC) and higher-order factor scales when adapted for use with archival adolescent MMPI data. *Archives of Assessment Psychology*, 3(1), 23–36.
- Tsai, D. C., & Pike, P. L. (2000). Effects of acculturation on the MMPI-2 scores of Asian American students. *Journal of Personality Assessment*, 74(2), 216–230.
- Tsushima, W. T., & Tsushima, V. G. (2009). Comparison of MMPI-2 validity scales among compensation-seeking Caucasian and Asian American medical patients. *Assessment*, 16(2), 159–164.
- Tucker, R. K., Weaver, R. L., Duran, R. L., & Redden, E. M. (1983). Criterion-related validity of three measures of assertiveness. *Psychological Record*, 33, 361–370.
- Turk, D., & Salovey, P. (1985). Cognitive structures, cognitive processes, and cognitive behavior modification. *Cognitive Therapy and Research*, 9, 19–33.
- Turkat, I. D. (1990). *The personality disorders: A psychological approach to clinical management*. New York, NY: Pergamon Press.
- Turner, S. M., DeMers, S. T., Fox, H. R., & Reed, G. M. (2001). APA's guidelines for test user qualifications. *American Psychologist*, 56, 1099–1113.
- Twentyman, C. T., & McFall, R. M. (1975). Behavioral training of social skills in shy males. *Journal of Consulting and Clinical Psychology*, 53, 393–401.
- Ullman, L. P., & Krasner, L. A. (1965). *Case studies in behavior modification*. New York, NY: Holt, Rinehart and Winston.
- Ulrich, L. P., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100–116.

- United States Census Bureau. (2015). *USA quickfacts*. Retrieved from <http://quickfacts.census.gov/qfd/states/00000.html>.
- Unoka, Z., Rózsa, S., Kő, N., Kállai, J., Fábíán, Á., & Simon, L. (2004). Validity and reliability of the SCL-90 in a Hungarian population sample. *Psychiatria Hungarica*, 19, 235–243.
- Urbán, R., Kun, B., Farkas, J., Paski, B., Kökönyi, G., Unoka, Z., et al. (2014). Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. *Psychiatry Research*, 216(1), 146–154.
- Urgelles, J. (2014). Assessing personality using self-report measures with African American clients. In L. T. Benuto & B. D. Leany (Eds.), *Guide to psychological assessment with African Americans* (pp. 43–62). New York, NY: Springer.
- Vanderploeg, R. D. (Ed.). (1999). *Clinicians' guide to neuropsychological assessment* (2nd ed.). New York, NY: Psychology Press.
- Vanderploeg, R. D., Schinka, J. A., Baum, K. M., Tremont, G., & Mittenberg, W. (1998). WISC-III premorbid prediction strategies: Demographic and best performance approaches. *Psychological Assessment*, 10, 277–284.
- Vandiver, T., & Sheer, K. J. (1991). Temporal stability of the diagnostic interview schedule. *Psychological Assessment*, 3, 277–281.
- Vane, J. R., & Guarnaccia, V. J. (1989). Personality theory and personality assessment measures: How helpful to the clinician? *Journal of Clinical Psychology*, 45, 5–19.
- Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of the Beck Depression Inventory-II: An evaluation. *Assessment*, 15, 177–187.
- Van Zomeren, A. H., & Brouwer, W. H. (1990). Attentional deficits after closed head injury. In B. G. Deelman, R. J. Saan, & A. H. Van Zomeren (Eds.), *Traumatic brain injury: Clinical, social, and rehabilitation aspects*. Amsterdam, The Netherlands: Swets & Zeitlinger.
- Vassend, O., & Skrandal, A. (1999). The problem of structural indeterminacy in multidimensional symptom report instruments: The case of the SCL-90-R. *Behaviour Research and Therapy*, 37, 685–701.
- Veiel, H. O. F., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, 13, 356–368.
- Vernon, P. E. (1964). *Personality assessment: A critical survey*. London, UK: Methuen.
- Vieta, E., Colum, F., Martinez-Aran, A., Benabarre, A., Reinares, M., & Gasto, C. (2000). Bipolar II disorder and comorbidity. *Comprehensive Psychiatry*, 41, 339–343.
- Viglione, D. J., Blume-Marcovici, A. C., Miller, H. L., Giromini, L., & Meyer, G. (2012). An inter-rater reliability study for the Rorschach Performance Assessment System. *Journal of Personality Assessment*, 94(6), 607–612.
- Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment*, 13, 452–471.
- Viglione, D. J., & Taylor, N. (2003). Empirical support for interrater reliability of Rorschach Comprehensive System coding. *Journal of Clinical Psychology*, 59(1), 111–121.
- Vigneau, F., & Cormier, S. (2008). The factor structure of the State-Trait Anxiety Inventory: An alternative view. *Journal of Personality Assessment*, 90, 280–285.
- Vincent, K. R., Castillo, I., Hauser, R., Stuart, H. J., Zapata, J. A., Cohn, C. K., & O'Shanick, G. J. (1983). MMPI code types and DSM-III diagnosis. *Journal of Clinical Psychology*, 39, 829–842.
- Vincent, K. R., & Harman, M. J. (1991). The Exner Rorschach: An analysis of its clinical validity. *Journal of Clinical Psychology*, 47, 596–599.
- Visser, R. S. H. (1992). *Manual for the Complex Figure Test*. Amsterdam, the Netherlands: Swets & Zeitlinger.
- Vogt, A. T., & Heaton, R. L. (1977). Comparison of WAIS indices of cerebral dysfunction. *Perceptual and Motor Skills*, 45, 607–615.
- Volker, M. A., Lopata, C., Vujnovic, R. K., Smerbeck, A. M., Toomey, J. A., Rodgers, J. D., ... Thomeer, M. L. (2010). Comparison of the Bender Gestalt-II and VMI-V in samples of typical children and children with high-functioning autism spectrum disorders. *Journal of Psychoeducational Assessment*, 28(3), 187–200.
- Wagner, R. K. (1949). The employment interview: A critical review. *Personnel Psychology*, 2, 17–46.

- Wagner, R. K. (1997). Intelligence, training, and employment. *American Psychologist*, 52, 1059–1069.
- Walsh, K. (1994). Neuropsychological assessment of patients with memory disorders. In S. Toyz, D. Byrne, & A. Gilandas (Eds.), *Neuropsychology in clinical practice* (pp. 107–127). New York, NY: Academic Press.
- Walton, R. G., & Elliott, D. (1980). *Residential care: A reader in current theory and practice*. New York, NY: Pergamon Press.
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York, NY: Routledge.
- Wang, X., Gao, L., Shinfuku, N., Zhang, H., Zhao, C., & Shen, Y. (2000). Longitudinal study of earthquake-related PTSD in a randomly selected community sample in north China. *American Journal of Psychiatry*, 157, 1260–1266.
- Ward, L. C. (1991). A comparison of the MMPI and the MMPI-2. *Psychological Assessment*, 3, 688–690.
- Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory–II. *Psychological Assessment*, 18, 81–88.
- Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment*, 24(2), 328–340.
- Ward, L. C., Ryan, J. J., & Axelrod, B. N. (2000). Confirmatory factor analysis of the WAIS-III standardization data. *Psychological Assessment*, 12, 341–345.
- Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools*, 32, 267–276.
- Warren, J. I., Stein, J. A., & Grella, C. E. (2007). Role of social support and self-efficacy in treatment outcomes among clients with co-occurring disorders. *Drug and Alcohol Dependence*, 89, 267–274.
- Watkins, C. E. (1991). What have surveys taught us about the teaching and practice of psychological assessment. *Journal of Personality Assessment*, 56, 426–437.
- Watkins, C. E., Campbell, V. L., & McGregor, P. (1990). What types of psychological tests do behavioral (and other) counseling psychologists use? *Behavior Therapist*, 13, 115–117.
- Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, 26, 54–60.
- Watkins, M. W., Glutting, J. J., & Lei, P. (2007). Validity of the Full Scale IQ when there is significant variability among the WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, 14, 13–20.
- Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33, 448–457.
- Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1966). *Pragmatics of human communication*. New York, NY: Norton.
- Way, B. B., & Banks, S. (2001). Clinical factors related to admission and release decisions in psychiatric emergency services. *Psychiatric Services*, 52(2), 214–218.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR–20: Assessing risk for violence (Version 2)*. Vancouver, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, 19, 87–95.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: Williams & Wilkins.
- Wechsler, D. (1974). *Wechsler Memory Scale manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1987). *Wechsler Memory Scale–Revised manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *The Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

- Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997a). *WAIS-III Administration and Scoring manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler Memory Scale: Administration and Scoring manual* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002a). *WAIS-III/WMS-III technical manual, updated*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002b). *Wechsler memory scale—Third Edition, abbreviated*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002c). *Wechsler Preschool and Primary Scale of Intelligence—3rd Ed.* San Antonio, TX: Harcourt Assessment.
- Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children—4th Ed.: Administration and scoring manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children—4th Ed.: Technical and interpretive manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale—4th Ed.: Administration and scoring manual*. San Antonio, TX: NCS Pearson, Inc.
- Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale—4th Ed.: Technical and interpretive manual*. San Antonio, TX: NCS Pearson, Inc.
- Wechsler, D. (2009a). *Wechsler Memory Scale—Fourth Edition (WMS-IV): Administration and scoring manual*. San Antonio, TX: NCS Pearson.
- Wechsler, D. (2009b). *WMS-IV technical and interpretive manual*. San Antonio, TX: NCS Pearson.
- Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). San Antonio, TX: NCS Pearson.
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th Ed.). San Antonio, TX: Harcourt Assessment.
- Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children—5th ed.: Administration and scoring manual*. San Antonio, TX: NCS Pearson.
- Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children—5th ed.: Technical and interpretive manual*. San Antonio, TX: NCS Pearson.
- Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (2004). *WISC-IV Integrated: Technical and research manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Harcourt Assessment.
- Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology*, 4, 233–265.
- Weed, L. L. (1968). Medical records that guide and teach. *New England Journal of Medicine*, 278, 593–600.
- Weed, N. C., Butcher, J. N., McKenna, T., & Ben-Porath, Y. S. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2: The APS and AAS. *Journal of Personality Assessment*, 58, 389–404.
- Weed, N. C., Butcher, J. N., & Williams, C. L. (1994). Development of MMPI-A Alcohol/Drug problem scales. *Journal of Studies on Alcohol*, 55, 296–302.
- Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, 44, 98–104.
- Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment*, 8, 206–213.
- Weiner, I. B. (1999). Incremental validity of the Rorschach. *Assessment*, 6, 327–338.
- Weiner, I. B. (2004). Rorschach assessment: Current status. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 343–355). Hoboken, NJ: Wiley.
- Weiner, I. B., & Exner, J. E. (1991). Rorschach changes in long-term and short-term psychotherapy. *Journal of Personality Assessment*, 56, 453–465.
- Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment*, 31(2), 114–131.

- Weiss, L. G., & Price, L. (2002). An update on the factor structure of the WMS-III. www.psychcorp.com/catg/Wms/wmsfactor.html.
- Weiss, L. G., Prifitera, A., & Roid, G. H. (1993). The WISC-III and the fairness of predicting achievements across ethnic and gender groups. *Monograph Series of the Journal of Psychoeducational Assessment: Wechsler Intelligence Scale for Children, 3rd ed.* 35–42.
- Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2015). *WISC-V assessment and interpretation: Scientist-practitioner perspectives*. San Diego, CA: Elsevier.
- Weissman, A., & Beck, A. T. (1978, November). *Development and validation of the Dysfunctional Attitude Scale (DAS)*. Paper presented at the 12th annual meeting of the Association for the Advancement of Behavior Therapy, Chicago, IL.
- Weller, R. A., Penick, E. C., Powell, B. J., Othmer, E., Rice, A. S., & Kent, T. A. (1985). Agreement between two structured psychiatric diagnostic interviews: DIS and the PDI. *Comprehensive Psychiatry*, 26, 157–163.
- Westcott, M. C., & Alfano, D. P. (2005). The Symptom Checklist–90–Revised and mild traumatic brain injury. *Brain Injury*, 19, 1261–1267.
- Westen, D., & Muderrisoglu, S. (2006). Clinical assessment of pathological personality traits. *American Journal of Psychiatry*, 163, 1285–1287.
- Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry*, 156, 258–272.
- Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, part II: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273–285.
- Wetter, M. W., Baer, R. A., Berry, D. T., Robinson, L. H., & Sumpter, J. (1993). MMPI-2 profiles of motivated fakers given specific symptom information: A comparison to matched patients. *Psychological Assessment*, 5, 317–323.
- Wetter, M. W., & Deitsch, S. E. (1996). Faking specific disorders and temporal response consistency on the MMPI-2. *Psychological Assessment*, 8, 39–47.
- Wetzler, S. (1990). The Millon Clinical Multiaxial Inventory (MCMI): A review. *Journal of Personality Assessment*, 55, 445–464.
- Wetzler, S., & Marlowe, D. (1990). “Faking bad” on the MMPI, MMPI-2 and Millon–II. *Psychological Reports*, 67, 1117–1118.
- Wetzler, S., & Marlowe, D. (1992). What they don’t tell you in the test manual: A response to Millon. *Journal of Counseling and Development*, 70, 427–428.
- Whisman, M. A., Judd, C. M., Whiteford, N. T., & Gelhorn, H. L. (2013). Measurement invariance of the Beck Depression Inventory—Second Edition (BDI–II) across gender, race, and ethnicity in college students. *Assessment*, 20(4), 419–428.
- Whitbeck, L. B. (2006). Some guiding assumptions and a theoretical model for developing culturally specific interventions with Native American people. *Journal of Community Psychology*, 34(2), 183–192.
- Whitcomb, S., & Merrell, K. W. (2012). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.
- White, G. W., Nielsen, L., & Prus, J. S. (1984). Head Start teacher and aide preferences for degree of specificity in written psychological recommendations. *Professional Psychology: Research and Practice*, 15, 785–790.
- White, S., & Edelstein, B. (1991). Behavioral assessment and investigatory interviewing. *Behavioral Assessment*, 13, 245–264.
- Whitelaw, S., Baldwin, S., Bunton, R., & Flynn, D. (2000). The status of evidence and outcomes in stages of change research. *Health Education Research*, 15, 707–718.
- Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the response bias scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology*, 23(7–8), 777–786.
- Whitworth, R. H., & McBlaine, D. D. (1993). Comparison of the MMPI and MMPI-2 administered to Anglo- and Hispanic-American university students. *Journal of Personality Assessment*, 61(1), 19–27.

- Wickramasekera, I. E. (1995a). A model of people at high risk to develop chronic stress-related somatic symptoms: Some predictions. *Professional Psychology: Research and Practice*, 17, 437–447.
- Wickramasekera, I. E. (1995b). Somatization: Concepts, data, and predictions from the high-risk model of threat perception. *Journal of Nervous and Mental Disease*, 186, 15–23.
- Widiger, T. A., & Clark, L. A. (2000). Toward DSM-V and the classification of psychopathology. *Psychological Bulletin*, 126, 946–963.
- Widiger, T. A., & Kelso, K. (1983). Psychodiagnosis of Axis II. *Clinical Psychology Review*, 3, 491–510.
- Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, 62(2), 71.
- Widiger, T. A., Williams, J., Spitzer, R., & Francis, A. (1985). The MCMI as a measure of DSM-III. *Journal of Personality Assessment*, 49, 366–378.
- Wiebe, J. S., & Penley, J. A. (2005). A psychometric comparison of the Beck Depression Inventory–II in English and Spanish. *Psychological Assessment*, 17, 481–485.
- Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R Third Factor: It's more than distractibility. *Psychological Assessment*, 2, 91–97.
- Wiener-Levy, D., & Exner, J. E. (1981). The Rorschach EA–ep variable as related to persistence in a task frustration situation under feedback conditions. *Journal of Personality Assessment*, 45, 118–124.
- Wiens, A. N. (1976). The assessment interview. In I. B. Weiner (Ed.), *Clinical methods in psychology* (pp. 3–60). New York, NY: Wiley.
- Wiesner, W., & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and the degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 67, 189–205.
- Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology*, 23(5), 419–427.
- Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs*, 80(22), 1–42.
- Wiggins, J. S. (1971). *Content scales: Basic data for scoring and interpretation*. Unpublished raw data.
- Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1(4), 305–316.
- Wilde, M. C. (2006). The validity of the Repeatable Battery of Neuropsychological Status in acute stroke. *Clinical Neuropsychologist*, 20, 702–715.
- Wilde, N. J., Strauss, E., Chelune, G. J., Hermann, B. P., Hunter, M., Loring, D. W., ... Sherman, E. (2003). Confirmatory factor analysis of the WMS-III in patients with temporal lobe epilepsy. *Psychological Assessment*, 15, 56–63.
- Willburger, E., Fussenegger, B., Moll, K., Wood, G., & Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learning and Individual Differences*, 18(2), 224–236.
- Wilke, C. M., Gold, J., Bartko, J. J., Dickerson, F., Fenton, W. S., Knable ... Buchanaan, R. W. (2002). Test-retest stability of the Repeatable Battery for the Assessment of Neuropsychological Status in schizophrenia. *American Journal of Psychiatry*, 159, 838–844.
- Wilke, C. M., Gold, J., Humber, K., Dickerson, F., Fenton, W. S., & Buchanan, R. W. (2004). Brief cognitive assessment in schizophrenia: Normative data for the Repeatable Battery for the Assessment of Neuropsychological Status. *Schizophrenia Research*, 70, 175–186.
- Wilkinson, G. S., & Robertson, G. J. (2007). *Wide Range Achievement Test–IV*. Wilmington, DE: Jastak.
- Williams, C. L., Butcher, J. N., Ben-Porath, Y. S., & Graham, J. R. (1992). *MMPI-A content scales: Assessing psychopathology in adolescents*. Minneapolis: University of Minnesota Press.
- Williams, J. E., & Weed, N. C. (2004). Review of computer-based test interpretation software for the MMPI-2. *Journal of Personality Assessment*, 83, 78–83.

- Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., ... Wittchen, H. U. (1992). The Structured Clinical Interview for DSM-III-R (SCID). II: Multisite test-retest reliability. *Archives of General Psychiatry*, 49, 630–636.
- Williams, J. M. (1991). *Memory assessment scales*. Odessa, FL: Psychological Assessment Resources.
- Williams, W. M. (2000). Perspectives on intelligence testing, affirmative action, and educational policy [Special series]. *Psychology, Public Policy, and Law*, 6 (1), 5–19.
- Williamson, D. A., Veron-Guidry, S., & Kiper, K. (1998). Assessment of health-related disorders. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 256–270). Boston, MA: Allyn & Bacon.
- Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. J. (1999). *Behavioral assessment of the dysexecutive syndrome manual*. Odessa, FL: Psychological Assessment Resources.
- Wilson, B. A., Cockburn, J., & Baddeley, A. (2003). *The Rivermead Behavioral Memory Test—II*. Gaylord, MI: National Rehabilitation Services.
- Wilson, T. E., & Evans, I. M. (1983). The reliability of target behavior selection in behavioral assessment. *Behavioral Assessment*, 5, 33–54.
- Wing, J. K., Cooper, J. E., & Sartorius, N. (1974). *Description and classification of psychiatric symptoms*. Cambridge, UK: Cambridge University Press.
- Witmer, J. M., Bornstein, A. V., & Dunham, R. M. (1971). The effects of verbal approval and disapproval upon the performance of third and fourth grade children of four subtests of the Wechsler Intelligence Scale for Children. *Journal of School Psychology*, 9, 347–356.
- Witt, J. C., & Elliott, S. N. (1983). Assessment in behavioral consultation: The initial interview. *School Psychology Review*, 12, 42–49.
- Wittman, C. L. M., & Koele, P. (1999). Explaining treatment decisions. *Psychotherapy Research*, 9, 100–114.
- Wittman, C. L. M., & van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment*, 23, 56–61.
- Wolber, G. J., & Carne, W. F. (1993). *Writing psychological reports: A guide for clinicians*. Sarasota, FL: Professional Resource Press.
- Wolfson, D. (1985). *Neuropsychological History Questionnaire*. Tucson, AZ: Reitan Neuropsychology Laboratory.
- Wolpe, J., & Lang, P. J. (1964). A fear survey schedule for use in behavior therapy. *Behavior Research and Therapy*, 2, 27–30.
- Wolpe, J., & Lang, P. J. (1969). *Fear Survey Schedule*. San Diego, CA: Educational and Industrial Testing Service.
- Wolpe, J., & Lang, P. J. (1977). *Manual for the Fear Survey Schedule*. San Diego, CA: EdITS.
- Wolpe, J., & Lazarus, A. (1966). *Behavior therapy techniques*. New York, NY: Pergamon Press.
- Wood, J. M., Garb, H. N., Nezworski, M. T., Lilienfeld, S. O., & Duke, M. C. (2015). A second look at the validity of widely used Rorschach indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013). *Psychological Bulletin*, 141(1), 236–249.
- Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment*, 6, 341–349.
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach Test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology*, 56, 395–430.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, 7(1), 3–10.
- Woodruff, R. A., Goodwin, D. W., & Guze, S. B. (1974). *Psychiatric diagnosis*. New York, NY: Oxford University Press.
- Woodworth, R. S. (1918). *Personal data sheet*. Chicago, IL: Stoelting.
- Woody, R. W. (Ed.). (1980). *Encyclopedia of clinical assessment* (Vol. 1). San Francisco, CA: Jossey-Bass.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical description and diagnostic guidelines*. Geneva, Switzerland: Author.

- Worthington, E. L., Hook, J. N., Davis, D. E., & McDaniel, M. A. (2011). Religion and spirituality. *Journal of Clinical Psychology, 67*(2), 204–214.
- Wright, A. J. (2010). *Conducting psychological assessment: A guide for practitioners*. Hoboken, NJ: Wiley.
- Wygant, D. B., Ben-Porath, Y. S., & Arbisi, P. A. (2004, May). *Development and initial validation of a scale to detect somatic over-reporting*. Paper presented at the 39th annual symposium on recent developments in the use of the MMPI-2 and MMPI-A, Minneapolis, MN.
- Wyndowe, J. (1987). The microcomputerized Diagnostic Interview Schedule: Clinical use in an out-patient setting. *Canadian Journal of Psychiatry, 32*, 93–99.
- Wysocki, J. J., & Sweet, J. J. (1985). Identification of brain-damaged, schizophrenic, and normal medical patients using a brief neuropsychological screening battery. *International Journal of Clinical Neuropsychology, 7*, 40–44.
- Yaloff, J., & Brabender, V. (2001). Ethical dilemmas in personality assessment courses: Using the classroom for in vivo training. *Journal of Personality Assessment, 77*, 203–213.
- Yang, J., Bagby, R. M., & Ryder, A. G. (2000). Response style and the revised NEO Personality Inventory: Validity scales and spousal ratings in a Chinese psychiatric sample. *Assessment, 7*(4), 389–402.
- Yang, J., McCrae, R. R., Costa, P. T., Jr., Dai, X., Yao, S., Cai, T., & Gao, B. (1999). Cross-cultural personality assessment in psychiatric populations: The NEO-PI-R in the People's Republic of China. *Psychological Assessment, 11*(3), 359–368.
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740–767.
- Yasui, M., & Dishion, T. J. (2007). The ethnic context of child and adolescent problem behavior: Implications for child and family interventions. *Clinical Child and Family Psychology Review, 10*(2), 137–179.
- Yates, B., & Taub, J. (2003). Assessing the costs, benefits, cost effectiveness, and cost benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15*, 478–495.
- Yoon, K., Schmidt, F., & Ilies, R. (2002). Cross-cultural construct validity of the five-factor model of personality among Korean employees. *Journal of Cross-Cultural Psychology, 33*(3), 217–235.
- Young, M. S., & Schinka, J. A. (2001). Research validity scales for the NEO-PI-R: Additional evidence for reliability and validity. *Journal of Personality Assessment, 76*(3), 412–420.
- Zabora, J. R., Smith-Wilson, R., Fetting, J. H., & Enterline, J. P. (1990). An efficient method for psychosocial screening of cancer patients. *Psychosomatics, 31*, 192–196.
- Zedeck, S., Tziner, A., & Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology, 36*, 355–370.
- Zilmer, E. A., Bell, J. D., Fowler, P. C., Newman, A. C., & Stutts, M. L. (1991). Wechsler Verbal-Performance I.Q. discrepancies among psychiatric inpatients: Implications for subtle neuropsychological dysfunctioning. *Archives of Clinical Neuropsychology, 6*, 61–71.
- Zimbardo, P. G. (1977). *Shyness: What it is and what to do about it*. Reading, MA: Addison-Wesley.
- Ziskin, J., & Faust, D. (2008). *Coping with psychiatric and psychological testimony* (4th ed.). Los Angeles, CA: Law and Psychology Press.
- Zlotnick, C., Shea, M. T., Pilkonis, P., Elkin, I., & Ryan, C. (1996). Gender dysfunctional attitudes, social support, life events, and depressive symptoms over naturalistic follow-up. *American Journal of Psychiatry, 153*, 1021–1027.
- Zolliker, A. (1943). Schwangerschaftsdepression und Rorschach'scher formdeutversuch. *Schweiz Archives Neurologie und Psychiatri, 53*, 62–78.
- Zotter, D. L., & Crowther, J. H. (1991). The role of cognitions in bulimia nervosa. *Cognitive Therapy and Research, 15*, 413–426.
- Zuckerman, E. L. (2003). *The paper office: Forms, guidelines, and resources to make your practice work ethically, legally, and profitably*. New York, NY: Guilford Press.
- Zuckerman, E. L. (2005). *The clinician's thesaurus three: A guidebook for wording psychological reports* (6th ed.). New York, NY: Guilford Press.

Author Index

A

- Aaroe, L. A., 120
Able, S., 684
Abramson, L. Y., 132
Abramson, W. G., 114
Achenbach, T. M., 47
Acker, M. B., 636
Ackerman, M. J., 46, 728, 729
Acklin, M. W., 518, 519, 524
Adamopoulou, A., 665
Adams, H. E., 116
Adams, K. M., 617
Adams, R., 648
Adams, R. L., 201
Adams, W., 629
Aegisdottir, S., 25, 30, 32, 72
Agbayewa, M. O., 659
Ahava, G. W., 663
Ahn, H., 674
Ahnberg, J. L., 662
Aiduk, R., 256
Aiken, L. R., 10
Ainsworth, M. D., 550
Ajdacic-Gross, V., 662
Akamine, T. X., 495
Akiskal, H. S., 106
Alamilla, S. G., 377
Albanese, A. L., 71
Albaugh, B., 260
Albert, S., 525
Alberti, L., 657
Alderman, N., 633
Aldwin, C. M., 659
Alfano, D. P., 658
Alfonso, V. C., 159, 204
Alimohamed, S., 672
Allard, G., 342
Allen, J. G., 663, 664
Allen, J. J., 61, 526, 527
Allen, R., 636
Allgaier, A. K., 796
Allik, J., 494
Allison, E., 683
Alloy, L. B., 114, 129–131, 695
Allyn, J. B., 724
Almagor, M., 254, 262
Al-Musawi, N.M.M., 665
Alnacs, R., 104
Alonso, R. R., 119–120
Alp, I. E., 520
Alterman, A. I., 373–374, 377
Altman, H., 94
Alvarez, H. K., 117
Amabile, T. M., 209
Ambrosini, P. J., 106, 107
Ambwani, S., 377, 492
Ames, L. B., 546, 550, 551, 553, 557, 563, 564
Ametrano, R. M., 702
Amieva, H., 669
Amin, K., 620
Anastopolis, A. D., 205
Ancona, M., 654, 665
Anderson, E. M., 674
Anderson, K. E., 136
Anderson, L. A., 25

- Anderson, S. J., 659
 Anderson, T. J., 521
 Anderson, V., 173
 Andreas, S., 657, 674
 Andreoli, A., 427–428
 Andrews, D. A., 81
 Andrews, H. B., 674
 Anestis, J. C., 255
 Anger, W. K., 73
 Angleitner, A., 491, 494, 496
 Anglin, D., 683–684
 Angold, A., 107
 Angst, J., 662
 Anker, M. G., 653
 Antonuccio, D. O., 675, 687
 Antony, M. M., 81, 85, 653, 668, 675, 683
 Aplin, J. D., 668
 Appel, L., 217, 527
 Appelbaum, M. I., 156
 Apple, C., 217
 Applebaum, P. S., 45
 Apter, A., 107
 Aram, D. M., 197
 Arb, J. D., 173
 Arbisi, P. A., 254, 255, 277, 347
 Archer, R. P., 7, 44, 139, 150, 243, 252, 256, 257, 259, 263, 273, 304, 334–335, 337, 342, 417, 513, 519, 521, 524, 525, 579
 Areane, P., 702
 Arganza, G. F., 55–56, 82
 Arizmendi, T. G., 672, 677
 Armbuster, G. L., 541, 558–559, 565
 Armengol, C. G., 58, 731, 742, 743
 Arnau, R. C., 663–665
 Arnkoff, D. B., 702
 Arnold, B., 61
 Arredondo, R., 340
 Arthur, W., Jr., 82, 83
 Arvey, R. D., 82, 83
 Asaad, G., 41–42
 Ascaso, E. E., 111
 Aschenbach, T. M., 107
 Ashton, K., 255
 Atkinson, D. R., 61
 Atkinson, L., 520
 Aue, T., 136
 Austin, A., 702
 Avison, W. R., 659
 Ax, A. F., 136
 Axelrod, B. N., 150–151, 202, 218, 225
 Ayuso-Mateos, J. L., 665
 Azarbad, L., 373–374
 Azocar, F., 702

B
 Baade, L. E., 200
 Bacchiochi, J., 257
 Bachman, J., 42
 Baddeley, A. D., 624, 627, 629
 Baer, R. A., 273, 279, 497
 Bagby, M., 255
 Bagby, R. M., 255, 257, 273, 347, 436, 437, 494, 496, 497, 664
 Baggaley, M., 169
 Bagge, C. L., 665
 Bahn, A. K., 79
 Bailey, J. M., 520
 Bailie, J. M., 649
 Bain, E., 494
 Baity, M. R., 579, 728
 Baker, J., 42
 Bakker, C., 133
 Bakker-Rabdau, M., 133
 Baldwin, A., 156
 Baldwin, C., 156
 Baldwin, S., 156, 700
 Balgopal, R., 71
 Ball, J. D., 521, 525, 579, 663
 Balla, D. A., 47
 Ballenger, J., 273, 497
 Bamgbose, O., 6, 42
 Bandura, A., 129, 133, 134
 Banks, S., 678

- Bannatyne, A., 187, 192, 205
 Bansal, A., 65
 Barak, O., 658
 Barber, T. X., 51
 Barberger-Gateau, O., 203
 Barefoot, J. C., 492
 Barkham, M., 675
 Barley, D. E., 672, 674
 Barlow, D., 42, 81, 653, 675, 683, 741, 752
 Barlow, D. H., 102, 707
 Barnett, D. W., 152
 Barnum, P. T., 718
 Bar-On, R., 153
 Barona, A., 201
 Barr, W. B., 3, 8, 215, 629
 Barrera, M., 702
 Barrios, F. X., 664, 665
 Barth, J., 618
 Barth, J. T., 616
 Barthlow, D. L., 254, 257, 329
 Bartholomew, D. J., 139
 Bartko, J. J., 650
 Bartlett, J. A., 110
 Basham, R. B., 259, 695
 Basso, M. R., 231, 257
 Bateman, A. W., 604
 Baucom, D. H., 133
 Bauer, R. M., 618
 Baum, K. M., 202
 Baum, L. J., 256
 Bauman, L. J., 668
 Baxter, D., 518, 519
 Bearse, J., 57, 58
 Beatty, W. W., 648, 649
 Beavin, J. H., 78
 Bech, P., 659
 Beck, A. T., 67, 104, 131, 132, 662–666
 Beck, J. G., 68, 118–119, 675
 Beck, S. J., 515, 528, 553–555, 557, 563, 564, 566, 574
 Becker, C. E., 620
 Becker, E. S., 665
 Becker, K. D., 683
 Beebe, D. W., 152, 177–178
 Beehr, T. A., 83
 Beekman, A.T.F., 683–684
 Behnke, S., 59
 Beidas, R. S., 702
 Beier, E. G., 79
 Beldia, G., 668
 Bell, J. D., 198–199
 Bell, R., 173
 Bellack, A. S., 81, 116, 119, 121, 131
 Bellenger, J. C., 104
 Bellow, L. L., 659
 Belter, R. W., 634
 Beltrán, M. D., 665
 Benabarre, A., 106
 Bender, L., 618, 633
 Benedict, K. B., 438, 439
 Benjamin, A. B., 107, 658
 Benn, A. F., 525
 Ben-Porath, Y. S., 19, 62, 72, 243, 248–250, 252–256, 258, 260, 261, 265, 277, 301, 329, 334, 335, 340, 342, 347, 353, 366, 494
 Bensaheb, A., 129
 Benson, N. F., 490
 Bentley, K. H., 675
 Benton, A. L., 203–204, 626, 629
 Berchick, R. J., 104
 Berg, C. J., 741
 Bergan, J., 693, 694, 697, 698
 Berger, P., 427–428
 Bergin, A. E., 251
 Bergman, M. A., 182
 Bergua, V., 669
 Bernal, G., 524, 680
 Bernard, B. A., 638
 Bernat, E., 524
 Bernert, R. A., 740–741
 Bernstein, J., 185

- Berry, D. T., 273, 279, 497
 Bertoni, M., 676
 Berzins, J. I., 676
 Besing, S., 133
 Beutler, L. E., 10, 31, 34, 68, 71, 81, 90,
 102, 655, 672, 675–677, 681, 683,
 684, 688, 690, 693, 694, 697–699,
 703, 704, 719, 741, 752
 Bevington, D., 683
 Bharucha, K. J., 649
 Bieling, P. J., 668
 Bienvenu, O. J., 106
 Bieber, J. L., 132
 Bigler, E. D., 8, 43, 139, 215, 217, 626,
 708
 Bilder, R. M., 631
 Binder, L., 73
 Binet, A., 139, 156, 415
 Bingham, R. P., 702
 Birchler, G. R., 117, 133
 Bissada, H., 373–374
 Bivens, A., 425
 Bjornsen, C. A., 493
 Black, F. W., 638
 Black, J., 493
 Blagys, M. D., 728
 Blair, G., 260
 Blais, M. A., 373–374, 438, 439, 715
 Blake, H., 649
 Blakey, W. A., 159
 Blalock, L. D., 188
 Blanchard, E. B., 687
 Blashfield, R. K., 257, 473
 Blatt, S. J., 665
 Blatter, D. D., 217
 Blau, K., 694
 Blazer, D. G., 109
 Bleich, A., 107
 Block, A. R., 255
 Blonigen, D. M., 493
 Bloom, C., 689
 Blount, A., 10, 42
 Blue, S., 119
 Blum, N., 741
 Blume-Marcovici, A. C., 519
 Boccaccini, M. T., 243, 377
 Bockian, N., 42
 Bode, R., 649
 Bodenlos, J., 654, 665
 Boerger, A. R., 335
 Boes, J. L., 657
 Bogojevic, G., 658
 Bohart, A. C., 672
 Bolinsky, P. K., 259, 342
 Boll, T. J., 199, 273
 Bollini, A. M., 42
 Bolton, B., 638
 Bombel, G., 517
 Bonarius, H., 31
 Bongar, B., 34, 71, 81, 655, 672
 Bonnie, R. J., 45
 Boodoo, G., 21, 64, 139
 Boone, D., 373–374
 Boone, K. B., 241, 617
 Borneman, M. J., 64
 Bornstein, A. V., 51
 Bornstein, R. A., 197, 217
 Bornstein, R. F., 423, 427, 431, 520, 559
 Borthwick-Duffy, S., 207
 Borus, J. F., 42, 104, 257
 Bossé, R., 659
 Boswell, J. F., 653
 Botella, L., 704
 Botwinick, J., 636
 Bouchard, T. J., 21, 64, 139
 Bouisson, J., 669
 Bowers, D. A., 223
 Bowler, R. M., 620
 Boykin, A. W., 21, 64, 139
 Bracken, B. A., 64–66
 Bradbury, T. N., 133, 201
 Bradley, V., 207
 Brady, C., 260

- Braff, D., 551
 Bramson, R., 663
 Brandenburg, N. A., 426
 Branet, E., 620
 Brannigan, G., 630, 635–641, 643
 Brantley, P. J., 654, 665
 Bratu, M. E., 432
 Braukhaus, C., 657
 Brawman-Mintzer, O., 104
 Brayfield, A. H., 53
 Brehm, J. W., 697
 Brehm, S. S., 697
 Breit, S., 133
 Breitbach, L., 133
 Breitholtz, E., 114, 130, 132
 Brekke, J. S., 657
 Brennan, K. A., 493
 Brenner, E., 708, 716, 726, 742
 Brewin, C. R., 130
 Brickell, T. A., 202
 Bricklin, H., 46
 Briskin, G. J., 634
 Brockington, I. F., 108
 Brod, M., 684
 Brodsky, S. L., 243
 Brody, N., 21, 64, 139
 Bronfenbrenner, U., 696
 Brooks, B. L., 229
 Brooks, R., 628
 Broshek, D. K., 616
 Brouwer, D., 625, 664
 Brouwer, W. H., 625
 Brown, C. N., 161
 Brown, G., 104
 Brown, G. K., 67, 662
 Brown, L., 81, 83, 155
 Brown, M. C., 729
 Brown, R. S., 428–429
 Brown, T. A., 675, 741
 Brtek, M. D., 83–84
 Bruce, T. J., 34, 68
 Brungardt, T. M., 148
 Brunnell-Neuleib, S., 669
 Brunner, N. A., 635
 Brunner, T. M., 669
 Bryant, E. L., 553
 Buchanan, R. W., 650
 Buchanan, T., 57, 58, 72, 73
 Buchheim, P., 427–428
 Bucholz, K. K., 107
 Budd, M. A., 41–42
 Buffington-Vollum, J. K., 7, 44, 139, 243, 375, 417
 Bufka, L. F., 654
 Buhler, C., 558
 Bühler, J., 664
 Buis, T., 273
 Bullard, K. M., 82
 Buntinx, W.H.E., 207
 Bunton, R., 700
 Burgess, P., 633
 Burke, M. J., 438
 Burlingame, G. M., 655
 Burnett, C. K., 133
 Burt, D. B., 623, 628, 629
 Burton, L. A., 3, 8, 215, 629
 Butcher, J. N., 19, 72–74, 243, 248, 248–249, 249–258, 261–263, 265, 272, 279, 301, 303, 326, 334, 337, 340, 655, 681, 697
 Butters, N., 620

C
 Cabral, R. R., 677
 Cacciola, J. S., 104
 Cacioppo, J. T., 132, 136
 Cahill, B. S., 665
 Cai, G., 108, 494, 497
 Cai, T., 494
 Caldwell, A., 260
 Caldwell, A. B., 335
 Caldwell, C., 82
 Caldwell-Andrews, A., 497

- Calestro, K., 259
 Calhoun, K. S., 116
 Calhoun, S. L., 186, 205
 Callaghan, G. M., 653
 Callahan, J. L., 702
 Callender, J. C., 82
 Calvin, J., 335
 Camara, W. J., 6, 7, 73, 139, 215, 243,
 417, 513, 615, 627, 629, 630, 634,
 663
 Campbell, K. A., 73
 Campbell, L. F., 679, 742
 Campbell, V. L., 3, 6, 73, 114, 139, 513
 Campbell, W., 110, 111
 Campion, J. E., 82, 83
 Campos, R. C., 665
 Canetti, L., 656
 Canivez, G. L., 151
 Cannell, C. F., 83
 Canyock, E. M., 241
 Caparrós-Caparrós, B., 659
 Caputi, P., 153
 Carey, S., 687
 Carkhuff, R. R., 91
 Carkuff, R. R., 78
 Carlini, M., 628
 Carlozzi, N. E., 649
 Carlson, J. F., 10, 68, 272
 Carne, W. F., 719
 Caron, C., 104
 Carpenter, D., 654
 Carroll, L. M., 188
 Carswell, L. M., 200
 Cartwright, R. D., 134
 Casacalenda, N., 687
 Casey, J. E., 648
 Casey, P., 665
 Cashel, M. L., 7, 383
 Cashiel, M., 105
 Caslyn, D. A., 60
 Caspi, A., 81
 Castillo, I., 257, 521
 Castillo, M. T., 219
 Castlebury, F. D., 330
 Castonguay, L. G., 653, 684, 688, 698
 Cauffman, E., 45
 Cavanagh, K., 679, 742–743
 Ceci, S. J., 21, 64, 139
 Cella, D., 212
 Cerny, J. A., 675
 Chae, J., 494
 Chahal, R., 108, 109
 Chakraborty, D., 273
 Chambers, W. J., 106, 107
 Chan, A. Y. C., 153
 Chan, D., 518
 Chance, J., 347
 Chang, H., 665
 Channabasavanna, S. M., 427–428
 Chaparro, N., 111
 Chapman, C., 129
 Charter, R. A., 437
 Chastain, R. L., 140, 201
 Chavez Balderas, R. A., 659
 Chelune, G. J., 217
 Chen, H., 150–151, 154
 Chen, L., 108
 Cheung, F. M., 248–249, 377, 495
 Cheung, S. F., 377
 Chiles, J. A., 42
 Chinot, L., 649
 Chiros, C. E., 664
 Chisholm, S. M., 254
 Cho, D. W., 94
 Choca, J. P., 425, 429, 431, 432, 636, 689,
 719
 Choi, E. J., 217
 Chojnacki, J. T., 256
 Choney, S. K., 260
 Choquette, K. A., 657
 Chor, K. H. B., 678
 Chorpita, B. F., 683

- Christiano, B. A., 132
 Chui, H. C., 625
 Church, A. T., 262, 377, 495
 Churchard, M., 111
 Ciarlo, J. A., 654
 Ciarrochi, J. V., 153
 Cicchetti, D. V., 11, 47
 Ciminero, A. R., 116
 Claassen, C. A., 58
 Clare, I.C.H., 758
 Clark, C., 620
 Clark, C. G., 260
 Clark, L. A., 102, 493
 Clark, M. E., 329
 Clarkin, I. F., 34, 71, 81, 655, 672, 677, 683, 688, 694, 698, 703
 Clarkin, J. F., 493
 Cliche, D., 104
 Clinton, D., 655
 Cloninger, C. R., 110
 Clopton, J. R., 263
 Coan, R., 553
 Cobb, J., 67
 Cochran, C. D., 656
 Cockburn, J., 629
 Cockell, S. J., 700
 Cofer, C. N., 347
 Cohen, J., 81
 Cohen, J. B., 554
 Cohen, N. J., 197, 219
 Cohen-Kettenis, P. T., 658
 Cohn, C. K., 257, 521
 Coid, J., 741
 Colby, F., 273–274
 Colligan, R. C., 250
 Collins, F., 687
 Collins, J. F., 687
 Collum, C. M., 226
 Colum, F., 106
 Compton, W., 107
 Condit, D. C., 241
 Cone, J. D., 22, 116–118
 Conklin, L. R., 675
 Connelly, B. S., 64
 Conners, C. K., 47, 625
 Conradt, J., 659
 Consoli, G., 628
 Constantino, M. J., 702
 Cook, A. C., 129
 Cook, L. P., 653
 Cook, R. S., 25
 Coolidge, F. L., 665
 Cooney, N. L., 108–109, 694
 Cooper, H. M., 51–52, 54, 118
 Cooper, J. E., 81
 Cooper, P. J., 132
 Cooper, W. H., 82–83, 554, 558–559
 Cooper, Z., 102
 Corbella, S., 704
 Corcoran, K., 10, 68
 Corkin, S., 620
 Cormier, L. S., 89
 Cormier, S., 669
 Cormier, W. H., 89
 Corrigan, J. D., 796
 Corruble, E., 623
 Corsica, J. A., 373–374
 Coryell, W. H., 106
 Costa, P. T., Jr., 252, 490–497
 Costello, A. J., 110
 Costello, E. J., 107, 110
 Cottler, L. B., 107, 109
 Cottrell, D., 8, 683
 Covi, L., 655–656
 Crabbe-Decleve, G., 547
 Crago, M., 672, 675, 677
 Craig, P. L., 201
 Craig, R. J., 423, 424, 429, 431, 432
 Craig, S. L., 702
 Craighead, W. E., 687
 Crary, W. G., 93, 94
 Craske, M. G., 675, 752

- Crawford, J. I., 654
 Crimmins, D. B., 129
 Cronshaw, S., 83
 Crost, N. W., 497
 Croughan, J. L., 8, 107
 Crowell, E. W., 205
 Crowther, J. H., 132, 254
 Cruise, K. R., 375
 Crumpton, E., 551
 Cruz Fuentes, C. S., 659
 Cucciare, M. A., 742
 Cuellar, I., 61
 Cullen, B.A.M., 106
 Cullen, K., 135
 Cummings, J., 203
 Cummings, N. A., 654
 Cunningham, T. R., 118
 Cyr, J. J., 520, 657, 658
- D**
- Dahlstrom, L. E., 252
 Dahlstrom, W. G., 53, 64, 243, 249, 252, 253, 256, 303, 697
 Dai, X., 494, 497
 Daldrup, R. J., 693, 694, 697, 698, 752
 Daleiden, E. L., 683
 Dalgard, O. S., 665
 Dam, H., 521
 Dana, R. H., 55–56, 59, 61, 62, 517, 526, 527, 549, 638
 Daniel, F. J., 273
 Daniel, M. H., 211
 Danton, W. G., 675
 Danzinger, W. L., 636
 Dao, T. K., 521, 577, 579
 Daseking, M., 154
 Dash, V. N., 208
 Davidson, H., 557, 567, 572
 Davies, M., 104, 107, 135, 257
 Davis, A., 68, 649, 677, 678, 704, 724, 726
 Davis, D. E., 702
 Davis, H. G., 340
 Davis, J. J., 255
 Davis, J. R., 636
 Davis, R., 422, 426
 Davis, R. D., 421, 432, 436
 Davis, W. W., 9, 116
 de Cuypere, G., 658
 de la Rubia, J. M., 665
 de Lima, M. P., 494
 Deardorff, W. W., 329
 Debord, J. B., 699
 Decker, S. L., 630, 635–641, 643
 Deeg, D.J.H., 683–684
 DeFruyt, F., 493, 494
 Deisinger, J. A., 374
 Deitsch, S. E., 273
 DeKoninck, J. M., 547
 Dekosky, S. T., 203
 Del Greco, L., 133
 Del Pilar, G. H., 494
 del Rio, C., 421
 D'Elia, L. F., 617
 Delis, D. C., 141, 146, 148, 171, 203–204, 618, 633
 Dellapietra, L., 241
 Demakis, G. J., 277
 DeMann, J. J., 185
 Dematteo, D., 44
 Dembroski, T. M., 493
 DeMers, S. T., 5, 55, 74
 Demidenko, N., 373–374
 den Boer, P. C., 679
 DeNelsky, G. Y., 675
 De-Nour, A. K., 656
 Derogatis, L. R., 655–658, 660, 685
 DeRubeis, R. J., 687
 Desmarais, S., 119
 Desmet, M., 664
 DeSouza, C., 662
 desRosiers, G., 623
 Deutsch, G., 210
 Devins, N. P., 42

Diamant, C., 118
 Diamond, S. P., 623–624
 Dickerson, F., 650
 DiClemente, C. C., 699, 700
 Dies, R. R., 2, 718
 DiGuiseppe, R., 117
 Dikmen, S. S., 197
 Dillard, J. P., 696
 Dimeck, P. T., 218–219
 Dimond, M., 657
 DiNardo, P. A., 102, 741
 Dinning, D., 133
 Dishion, T. J., 683–684
 Dixon, L., 649
 Dobbins, C., 175, 199, 796, Russell
 Dobson, K. S., 662
 Dolle, K., 665
 Dolman, R., 218–219
 Domenech Rodriguez, M., 680
 Donaldson, J., 377
 Dong, Y. T., 262
 Donias, S., 659
 Donna, P. A., 725
 Donovan, P., 114
 Dorr, D., 432, 438
 Dougherty, T. W., 82
 Douglas, K. S., 45
 Dowd, E. T., 696, 698, 699
 Downs, W., 664
 Dowrick, C., 665
 Dozios, D. J., 662, 663
 Dozois, D. J., 664
 Drab, D. L., 700
 Draguns, J. G., 558, 559
 Driscoll, J., 152
 Dritschel, B. H., 132
 Drozdick, L. W., 154, 226–228
 Druss, B., 42
 DuAlba, L., 253–254
 Dubois, B., 203
 Dudek, S. Z., 549

Duff, K., 648–650
 Duhe, A., 133
 Duke, M. C., 517
 Dulcan, M. K., 109
 Dumitrascu, N., 517, 547
 Dumont, F., 7
 Duncan, B. L., 653
 Duncan, M. K., 109
 Duncan, S., 255
 Dungan, D. S., 634
 Dunham, R. M., 51
 Dunlop, B. W., 687
 Dunn, D. M., 626
 Dunn, G., 665
 Dunn, L. M., 626
 Dunn, T., 432
 Dura, J. R., 658, 663
 Duran, R. L., 120
 Durand, V. M., 129
 Durbin, J., 678
 Durham, R. L., 217
 Durham, T. W., 330
 Duthie, B., 420
 Dutra, R. L., 188
 Dutton, G. R., 654, 665
 Dwyer, E. K., 702
 Dyck, M. J., 129
 Dye, D. A., 492
 Dye, E., 109

E

Earls, R., 110
 Eaton, W. W., 108, 109
 Eaves, D., 45
 Ebert, R. J., 82
 Ebesutani, C., 683
 Ebner, D. L., 551
 Ebrahimkhani, N., 665
 Edelbrock, C. S., 109, 110
 Edell, W. S., 524
 Edelstein, B., 82

- Edens, J. F., 45, 375
 Eddins, G., 6, 9, 42, 653, 677
 Edwards, D. W., 256, 265, 303
 Edwards, G. L., 634
 Eells, T. D., 681
 Egeland, B. R., 52
 Ehler, J. G., 106
 Eich, D., 662
 Eidelman, B. H., 687
 Eidelson, R. J., 133
 Eisman, E. J., 3
 Ekelman, B. L., 197
 Ekstrom, D. E., 94
 Elashoff, J., 51
 Elhai, J. D., 163, 164
 Elion, V. H., 65
 Elkin, I., 687, 695
 Elklit, A., 425
 Ellens, B. M., 57, 58, 74, 273
 Elliot, A. J., 121
 Elliott, D., 677
 Elliott, R., 657, 672
 Elliott, S. N., 124
 Elwood, R. W., 217
 Embretson, S., 141
 Emery, G., 662
 Emmanuel, N., 104
 Emslie, H., 633
 Endicott, J., 81, 101, 104, 105
 Engel, R. R., 377
 Engeldal, K., 669
 Engle, D., 693, 694, 697, 698, 752
 English, L. T., 277
 Enterline, J. P., 659
 Eppinger, M. G., 201
 Epstein, J., 85
 Epstein, N., 133
 Erard, R. E., 514, 518
 Erb, A. L., 57
 Erdberg, P., 514, 516, 518, 523, 541, 550, 553–555, 559, 563, 571, 574, 579
 Erez, E., 256
 Erginel, A., 566
 Erickson, R. C., 620
 Ernst, D., 129
 Essau, C. A., 659
 Evans, D. L., 94
 Evans, E., 683–684
 Evans, I. M., 123
 Evans, J. J., 633
 Evenson, R. C., 94
 Exner, D. E., 515–516, 548, 549, 551, 553, 575
 Exner, J. E., 55, 516–522, 524, 525, 527–532, 534–536, 539, 541, 543, 545, 547, 549–556, 558–581, 587
 Eyde, L. D., 55, 514
 Eysenck, H. J., 673, 674
- F**
 Fábíán, Á., 659
 Fairbank, J. A., 428
 Fairburn, C. G., 102, 657, 658
 Falissard, B., 623
 Fallows, R. R., 181
 Fals-Stewart, W., 133, 375
 Fan, R. M., 377
 Fantoni-Salvator, P., 377
 Fantuzzo, J. W., 159, 160, 169
 Faravelli, L., 628
 Farkas, J., 658
 Farmer, A., 108
 Farmer, J., 132
 Farmer, R. F., 114, 116
 Faust, D., 6, 28, 29, 44, 73, 342
 Feighner, J. P., 101
 Fein, D., 141, 146, 171
 Feingold, A., 177
 Feldman, H. H., 203, 229
 Feldman, S. E., 51
 Fenton, W. S., 650
 Fernandez, K., 377
 Fernandez-Alvarez, H., 704

- Fetting, J. H., 659
Fichtenberg, N. L., 648
Fiel, A., 436
Field, K., 638
Fields, J. A., 628
Fillenbaum, G., 620
Filskov, S. B., 644
Finch, A. J., 634
Finch, W. H., 649
Fincham, F. D., 133
Fine, J. D., 150
Finger, M. S., 73
Fink, A. D., 551
Fink, J., 652
Finn, J. A., 255
Finn, S. E., 3, 6, 58, 272, 420, 726, 727, 729, 730, 737, 742, 743
Fiorello, C. A., 204
First, M. B., 9, 102–104, 116, 683, 741
Fischer, C. T., 272, 727, 730
Fischer, J., 10, 68
Fischhoff, B., 29
Fisher, D. C., 71, 197, 219, 655
Fisher, J. E., 672, 708
Fisher, P. W., 109
Fiske, D. W., 28–29, 204
Flanagan, D. P., 139, 147, 150, 153, 161, 166–168, 173
Flato, C. G., 377, 492
Flaugher, R. L., 64
Fleischmann, D. A., 623
Fleming, B., 133
Fleming, J., 681
Flemming, M. Z., 152
Fletcher-Jantzen, E., 65, 211
Flores, J., 373–374, 377
Florio, C. M., 520
Flynn, D., 700
Flynn, P. M., 428, 430
Foa, E. B., 675
Fode, K. L., 51
Follette, W. C., 9, 116, 677
Folstein, M. F., 94, 624
Folstein, S. E., 94, 108–109, 624
Fonagy, P., 604, 683
Fonda, C. P., 558
Fontaine, J., 664
Forbey, J. D., 9, 72, 249
Fornal, R. E., 93
Forrester, B., 676
Foster, S. L., 22
Fournier, J. P., 104
Fowler, J. C., 163, 164, 518, 519, 741
Fowler, P. C., 198–199
Fox, C. M., 657
Fox, D., 277
Fox, H. R., 5, 55, 74, 525
Fraboni, M., 252
Francis, A., 9, 116, 428
Frank, E., 104
Frank, G., 524, 561
Frank, J. B., 674
Frank, J. D., 674
Franke, G. H., 659
Franzen, M. D., 116, 273
Fraser, A. W., 51
Fray, P. J., 73
Free, M. L., 687
Frei, A., 662
Freilich, B. M., 649
French, J., 521, 525
Freyre, M. A., 665
Friedman, A. F., 106, 263, 342
Friedman, R. M., 110
Friend, R., 132
Friman, P. C., 109, 110
Frueh, B. C., 525, 663, 664
Fuchs, D., 51
Fuchs, L. S., 51
Fuld, P. A., 180
Fuller, A. K., 473
Furman, J.M.R., 188, 204

Furukawa, T. A., 664, 665

Fussenegger, B., 195

Fyer, A. J., 105

G

Gale, S. D., 217

Galina, H., 551

Gallagher, R. E., 139

Gallagher-Thompson, D., 674–676

Gallucci, N. T., 253–254

Gambrill, E. D., 133

Gamma, A., 662

Gan, Y., 495

Gao, B., 494, 497

Gao, L. K., 659

Garb, H. N., 6–8, 25, 28–30, 73, 81, 82, 85, 254, 517, 518, 520, 521

Garbin, M., 663

Garcia, C. G., 659

Gardner, H., 141, 153

Garfield, S. L., 132, 675

Garfinkel, P. E., 132

Garland, B. H., 377, 492

Garner, D. M., 132

Garratt, G., 129–131

Garrick, T., 649

Garrison, B., 663

Garvey, M. J., 687

Gary, H. E., 624

Garyfallos, G., 665

Gass, C. S., 217, 729

Gasto, C., 106

Gaw, K. F., 690, 693, 697, 698

Geary, D. C., 568

Geffken, G. R., 743

Gega, L., 679, 742–743

Geisinger, K. F., 10, 62, 68, 272

Gelhorn, H. L., 665

Geller, J., 700

George, L., 678

George, L. K., 109

George, L. M., 528

Gerardi, M. A., 687

Gerstle, R. M., 568

Gervais, R. O., 347

Getter, H., 694

Ghassemzadeh, H., 665

Giannakou, M., 665

Gibbon, M., 8, 102–104, 257

Gibbons, R. D., 687

Gibertini, M., 426

Giffen, M. B., 574

Gilboa, A., 658

Gill, H. S., 551

Gill, M., 835

Gilley, D. W., 623

Gillis, J. R., 437

Gilmore, D. C., 83

Ginting, H., 665

Girdwood, C. P., 114

Giromini, L., 519, 527

Glaser, D., 683

Glass, C. R., 114, 117, 132, 702

Glass, D. R., 687

Glass, G. V., 674

Glass, L. A., 161

Glenn, W. J., 277

Gliner, J. A., 18

Glueckauf, R. L., 10, 42

Glutting, J. J., 152, 165, 169

Gocka, E., 334–335

Godber, T., 173

Goddard, R., 342

Godfrey, H.P.D., 619

Goering, P., 678

Goethe, K. E., 624

Goetz, C. G., 623

Gogos, A., 649

Goheen, M. D., 699

Gold, J., 650

Goldbeck, L., 154

- Goldberg, E., 631
 Goldberg, J., 437
 Goldberg, L. S., 119
 Goldberg, R., 649
 Golden, C. J., 8, 198, 617, 620
 Golden, R. M., 94
 Goldfrank, R. R., 42
 Goldfried, M. R., 116, 117, 120
 Goldman, D., 260
 Goldring, E., 8
 Goldstein, A. M., 43, 44, 617, 618
 Goncalves, A. A., 432
 Gonçalves, B., 665
 Gontkovsky, S. T., 648, 649
 González Castro, F. G., 702
 Goodglass, H., 203, 626
 Gooding, P., 677
 Goodman, G. M., 262
 Goodman, R., 649
 Goodstein, L. D., 25
 Goodwin, B. E., 255
 Goodwin, D. W., 101
 Goodwin, G. M., 623
 Gordon, N. G., 335
 Gorman, J. M., 674, 683
 Gorsuch, R. L., 159, 160, 666, 685
 Gorwood, P., 623, 628
 Gottesman, I. I., 259, 262, 640
 Gottfried, E. D., 255
 Gottheil, E., 655
 Gotts, E. E., 334–335
 Gough, H. G., 492
 Gould, R. A., 675
 Grace, J., 633
 Graham, J. R., 19, 64, 65, 243, 248, 249, 251, 253–257, 259–261, 263, 301, 329, 334–335, 340, 697
 Graham, P., 79
 Grant, I., 227, 617, 620
 Graves, N. S., 200, 201
 Graves, R. E., 658
 Grayson, D., 493
 Greaves, S. T., 528
 Grebstein, L., 663
 Green, A., 649
 Green, C. J., 7, 67
 Green, K., 201, 684
 Green, R. E., 201
 Green, S. B., 260
 Greenbaum, P. F., 110
 Greenberg, J., 657
 Greenberg, L., 131
 Greenberg, L. S., 672
 Greene, P., 347
 Greene, R., 65, 258
 Greene, R. L., 65, 248, 260–263, 265, 334, 335
 Greene, R. W., 117
 Greeno, C., 104
 Greenspan, S., 152
 Greenway, P., 206
 Gregory, C., 256
 Greiffenstein, M. F., 277
 Grella, C., 683–684
 Grella, C. E., 695
 Gresham, R. M., 121
 Grieger, I., 82
 Griffin, B., 493, 497
 Griffin, R., 256
 Grigorenko, E. L., 66, 156
 Grillo, J., 428–429
 Grimm, L. G., 80
 Grisso, E. P., 45
 Groen, G., 659
 Groenvynck, H., 664
 Grohol, J. M., 679, 742
 Gronnerod, C., 519
 Grossman, L. S., 525
 Grossman, S., 19, 67, 417, 418, 421
 Grossman, S. J., 73

- Grothe, K. B., 654, 665
 Groth-Marnat, G., 7–10, 32, 34, 42, 43, 67, 68, 73, 74, 81, 90, 139, 152, 171, 180, 188, 192, 196, 198, 206, 273, 518, 625, 653–655, 672, 673, 675, 677, 678, 683, 704, 708, 718, 719, 724, 726, 739, 741
 Grotzer, T. A., 156
 Grove, W. M., 7, 25, 30, 520
 Groves, J. A., 377
 Growdon, J. H., 620
 Gruber, C. P., 47
 Guarino, J., 131–132
 Guarnaccia, V. J., 31
 Gudjonsson, G. H., 758
 Guevremont, G. C., 114, 117, 121, 130
 Guilmette, T. J., 619
 Guldager, S., 425
 Gunderson, J., 657
 Gunderson, J. G., 497
 Gupta, A. S., 635, 649
 Gupta, N., 635, 649
 Gutierrez, P. M., 664, 665
 Gutkin, T. B., 154
 Guze, S. B., 101
 Guzman, A., 426
 Gwynne, J., 51
 Gynther, M. D., 260
 Gyoerkoe, K. L., 525
- H**
 Haaga, D. A., 129
 Haas, A. P., 90
 Haber, J. C., 256
 Hagaman, R., 677
 Hahn, J., 73, 272
 Hain, J. D., 636
 Hale, J. B., 139, 204
 Hale, W. D., 656
 Hales, R. E., 740–741
 Haley, E. M., 558
 Haley, W. E., 696
 Halfon, O., 649
 Hall, C., 73
 Hall, C. R., 218–219
 Hall, G.C.N., 65, 248–249, 260–262, 702
 Hall, J. E., 10
 Haller, D. L., 250
 Hallmark, R., 3, 6, 139, 513
 Halstead, W., 617
 Hamel, M., 523
 Hammeke, T. A., 8, 173, 616, 617, 708
 Hammeke, T. E., 708
 Hammen, C. L., 131
 Hammill, D. D., 155
 Hammond, J. A., 273
 Hamsher, K., 203–204, 626
 Han, K., 262, 279
 Handel, R. W., 7, 44, 62, 139, 243, 248–249, 258, 342, 417
 Handler, L., 272, 520, 526, 727
 Handwerk, M. L., 110
 Hanekamp, M., 104
 Hansell, A. G., 573
 Hansen, N. B., 655
 Hanson, D. R., 259
 Hanson, R. K., 45, 251, 520
 Hara, N., 683–684
 Harding, C. M., 684
 Hardy, G. E., 675
 Hargrave, G. E., 256
 Harioan, J., 523
 Harkness, A. R., 19, 250, 366
 Harmon, R. J., 18, 521, 561
 Harper, C., 649
 Harrell, T. H., 256
 Harring, J. R., 657
 Harrington, K., 231
 Harris, M. J., 273
 Harris, R. E., 295, 334
 Harrison, D. F., 119
 Harrison, D. W., 623
 Harrison, P. L., 139

- Hart, S. D., 45
Harte, D. K., 114, 134
Hartkamp, N., 658, 659
Hartlage, L., 64
Hartshorne, H., 115
Harvey, V. S., 726, 727
Harwood, T. M., 34, 68, 90, 672, 675,
676, 681, 684, 685, 688, 690, 693,
694, 698, 703, 719
Hashimoto, N., 664
Hasselhorn, M., 195
Hastie, R., 28, 29
Hatch, A. L., 42
Hatfield, D. R., 653
Hathaway, S. R., 243, 246, 247–248, 251,
338, 492
Hatt, C. V., 205
Hattem, D., 126
Hauff, E., 659
Hauser, R., 257, 521
Haverkamp, B. E., 28
Haviland, M. G., 664
Hawkins, K. A., 197, 217, 219, 234, 238
Hawkins, R. J., 48, 653
Hawkins, S. A., 28, 29
Haxby, J., 620
Hayes, J. A., 653, 657, 658
Hayes, S. C., 9, 116
Haynes, S. N., 82, 114–119, 121
Heaton, R. K., 64, 227, 617
Heaton, R. L., 196
Hebben, N., 739
Hebert, K. R., 182
Heckrath, L., 657
Hedeker, D., 687
Hedgepeth, B. E., 656
Hedlund, J. L., 94
Heiby, E. M., 114, 121
Heilbrun, A. B., Jr., 492
Heilburn, K., 43, 44
Heimberg, R. G., 119, 132
Heinberg, L. J., 255
Heinemann, A., 649
Heiser, L. J., 708, 740
Hellkamp, D. T., 636
Hellwig-Brida, S., 154
Helmes, E., 249
Helms, E., 252
Helms, J. L., 757, 758
Helzer, J. E., 8, 85, 101, 107–110
Hendin, H., 90
Henry, B., 81, 83, 130
Heraldsen, I. R., 658
Herjanic, B., 111
Herman, D. O., 211
Herman, S. M., 255
Hermann, B. P., 219
Hernández-Guzmán, L., 665
Hersen, M., 34, 81, 114–117, 119, 121,
131, 528, 668, 669, 708
Hertz, M. R., 515, 553, 555
Herzog, D. B., 105
Hesketh, B., 493
Hesse, M., 425
Heyman, A., 620
Heyne, L. K., 57
Hiatt, D., 256
Hidaka, M., 659
Higgins, R. L., 119–120
High, W. M., 624
Hill, J. S., 260
Hiller, J. B., 551, 559, 561, 574
Hillix, W. A., 52
Hillman, L., 571
Hilsabeck, R. C., 181, 428–429
Hilsenroth, M. J., 44, 330, 518, 519, 526,
527, 577, 579, 728
Himelstein, P., 568
Hiroe, T., 664
Hirsch, M., 107
Hishinuma, E. S., 669
Ho, R. M., 249
Hodgson, M. J., 188, 204

- Hoehn-Saric, R., 106
 Hofstede, G., 495
 Hogan, J., 20
 Hogan, L., 57, 58, 728
 Hogan, M. E., 114, 636
 Hogan, R., 20
 Hogarty, P. S., 156
 Hoge, R. D., 45, 81, 641, 643, 708, 719
 Hoge, S. K., 45
 Holdnack, J. A., 161, 201, 202, 225, 227, 228, 237
 Holland, A. L., 626
 Hollett, J., 81
 Hollis, J. W., 725
 Hollon, S. D., 130–132, 134, 687
 Holm Linneberg, I., 425
 Holman, J., 698
 Holmes, C. B., 634
 Holt, H., 676
 Holt, R. R., 550
 Holtzman, W. H., 522
 Hom, M. A., 740–741
 Honaker, L. M., 256
 Hood, M., 373–374
 Hoofien, D., 658
 Hook, J. N., 702
 Hope, D. A., 132
 Hopwood, C. J., 226, 377, 492
 Horner, M. D., 649
 Horowitz, M. J., 134
 Horvath, A. O., 698, 699
 Horvath, J. A., 141, 152, 153, 239
 Horvath, L. S., 708, 719
 Hostetler, K., 301
 Houts, A. C., 676, 741
 Houtsinger, J. K., 42
 Howes, M. J., 42
 Howieson, D. B., 8, 43, 139, 215, 626, 708
 Hoyt, M. F., 81
 Hser, Y., 683–684
 Huang, L. N., 55–56, 82
 Hubley, A. M., 94
 Huckleba, W., 180
 Huddleston, E. N., 334–335
 Huff, F. J., 620
 Huffcutt, a. I., 83
 Hufford, M. R., 81, 117
 Hughes, J. P., 620
 Humber, K., 650
 Humphrey, D. H., 256, 303
 Hunnicutt, L. C., 159
 Hunsley, J., 24, 68, 83, 251, 520
 Hunter, E. E., 662
 Hunter, J. E., 64, 83, 139, 140
 Hunter, M., 219
 Hutt, M. L., 634, 635
 Hwang, W. C., 702
 Hyer, L. A., 425, 649
 Hynan, L. S., 551
- I**
- Iacano, W. G., 136
 Iannone, C., 663
 Ilies, R., 494
 Imber, S. D., 687
 Imel, Z. E., 672, 674, 676
 Imhof, E. A., 250, 521, 525
 Ingersoll, J., 256, 257
 Ingram, R. E., 129, 131, 132
 Inoue, K., 684
 Irwing, P., 376
 Iverson, G. L., 154, 201, 227, 228, 273, 276
 Izat, Y., 796
- J**
- Jackson, D. D., 78
 Jackson, D. N., 252
 Jackson, J. S., 82
 Jackson, K. M., 375
 Jackson, R. L., 105
 Jacob, T., 493, 687
 Jacobs, G. A., 666, 685

Jacobson, J. M., 259, 337
 Jacobson, L., 62
 Jacobson, N. S., 677
 Jacova, C., 203
 Jaibert, A., 188
 Jakubowski, P., 366
 Jang, K. L., 491
 Jankowski, D., 432
 Janus, M., 259
 Jarvik, L. R., 180
 Jarvis, P. E., 618
 Jenkins, M., 551
 Jensen, A. L., 110
 Jensen, A. R., 62, 64, 618, 625
 Jensen-Doss, A., 683
 Jesse, R. C., 6
 Jewell, J., 257
 Johansson, B., 114
 John, R. S., 126
 Johnson, A., 159
 Johnson, B. N., 687
 Johnson, C. W., 93, 94
 Johnson, J., 551, 681
 Johnson, J. H., 252
 Johnson, J. W., 80
 Johnson, K. N., 252
 Johnson, L. C., 657, 659
 Johnson, M., 104
 Johnson, M. H., 106
 Johnson, R. L., 110
 Johnson, S. C., 217
 Johnson, S. K., 155
 Johnston, J., 684
 Joiner, T. E., 255
 Jones, C. H., 153
 Jones, G. N., 654, 665
 Jones, M. C., 115
 Jones, R. G., 132
 Jongsma, A. E., 34, 68
 Jonker, C., 684
 Jonson, J. L., 68, 272

Jordan, B. D., 649
 Jordan, N., 678
 Jorgensen, A. E., 521
 Joshua, N., 649
 Juan-Ferrer, J., 659
 Judd, C. M., 620, 665
 Judson, A. J., 347
 Jung, K. G., 110
 Jung, S., 226

K

Kabacoff, R. I., 668, 669
 Kablinger, A. S., 619
 Kadden, R. M., 108–109, 694
 Kadlubek, R. M., 153
 Kaemmer, B., 243, 249, 697
 Kagee, A., 665
 Kahn, M. W., 525
 Kahn, R. L., 83
 Kahn, T. C., 574
 Kalas, R., 109
 Kalembe, V., 273
 Kállai, J., 659
 Kallasmaa, T., 494
 Kalpakjian, C. Z., 219
 Kamin, L. J., 62
 Kamphaus, R. W., 7, 47, 634
 Kane, R. L., 73
 Kanfer, F. H., 80, 120, 122
 Kang, J., 104
 Kaplan, E., 10, 139, 141, 146, 148, 171,
 173, 177, 198, 199, 203, 204, 618,
 626, 633, 726
 Kaplan, R. M., 119, 154
 Kaplar, M. E., 116–117
 Karamghadiri, N., 665
 Karastergiou, A., 659
 Kareken, D. A., 29
 Karg, R. S., 103, 741
 Karr, C., 256
 Karson, M., 526
 Kasl-Godley, J., 696

- Kaslow, N. J., 42
 Kaspar, J. C., 635
 Kaster-Bundgaard, J., 121
 Kathol, R., 10, 42
 Katigbak, M. S., 495
 Katon, W. J., 41–42
 Katsikitis, M., 67
 Kaufman, A. S., 47, 65, 71, 139, 140,
 147, 150–153, 155, 161, 166–168,
 173, 174, 191, 192, 197–199,
 204–206, 209, 218, 225
 Kaufman, N. L., 47, 48, 206, 209
 Kawai, M., 665
 Kay, G. G., 9, 514
 Kazaoka, K., 549
 Kearney, C. A., 129
 Keeley, M. L., 743
 Kehl-Fie, K. A., 373–374
 Keiller, S., 255
 Keith, T. Z., 150, 151, 154
 Kelleher, K. J., 42
 Keller, F., 664
 Keller, M. B., 105, 106
 Keller, S. E., 110
 Kelley, C. K., 260
 Kellison, I., 743
 Kelly, E. L., 28–29, 663, 664
 Kelly, M. P., 217
 Kelly, S., 684
 Kelso, K., 420
 Kendall, P. C., 129–132, 702
 Kendjelic, E. M., 681
 Kennedy, B. L., 668
 Kent, R. N., 118
 Kent, T. A., 109
 Kerns, L. L., 83
 Keyser, D., 42
 Kiecolt-Glaser, J. K., 658, 663
 Killgore, W.E.S., 241
 Kim, B. S., 61
 Kim, H. K., 262
 Kim, S., 262
 Kimpara, S., 694
 Kincel, R. L., 549
 Kinder, B. N., 6, 496, 525
 King, L. C., 649
 King, N. J., 683
 Kinney, D. I., 649
 Kinoshita, Y., 664
 Kiosscoglou, G., 665
 Kiper, K., 116
 Kirk, S. A., 473
 Kirsch, I., 132
 Kiuse, J., 659
 Klassen, R. M., 130
 Kleespies, P. M., 696
 Klein, D. F., 105
 Klein, D. N., 107
 Klein, R. G., 526
 Kleinmuntz, B., 28–30
 Klinefelter, D., 250
 Klinetob, N., 260
 Klinkenberg, W. D., 85
 Klonoff, E. A., 61
 Klopfer, B., 515, 550, 553, 554, 557, 567,
 572
 Klopfer, W. G., 550, 718, 726–727, 738
 Klosko, J. S., 675
 Knaster, C. A., 260
 Knight, R. G., 619
 Knowles, E. S., 579
 Knudson, T. E., 334–335
 Kō, N., 659
 Koele, P., 688
 Koenen, K. C., 683–684
 Kohli, A., 635
 Kojima, M., 665
 Kōkōnyei, G., 658
 Konold, T. R., 152, 165, 169
 Koopman, R. F., 201, 202
 Kopelman, M. D., 627
 Kopper, B. A., 664, 665
 Koppitz, E. M., 634, 635, 645

Koren, D., 262
 Korenblum, M., 111
 Kosmides, M., 665
 Kostlan, A., 24, 254
 Kovera, C. A., 73
 Kramer, J., 146
 Kramer, J. H., 171, 180, 198–199,
 203–204, 618, 633
 Kramer, M., 79
 Krantz, S., 131
 Kranzler, J. H., 150
 Krasner, L. A., 115
 Kratochwill, T. R., 124
 Kraus, D. R., 653
 Krebs, P. M., 700
 Kremer, T., 496
 Kreukels, B.P.C., 658
 Krishnamurthy, R., 252, 334–335, 521
 Kristjansoon, B., 94
 Kroll, L. S., 273
 Krueger, R. F., 489, 491
 Kruse, J., 657
 Kubiszyn, T. W., 3
 Kubo, C., 659
 Kucharski, L. T., 255
 Kuhn, R., 558
 Kulhara, P., 635
 Kumar, G., 664
 Kun, B., 658
 Kuncze, J. T., 549
 Kurtz, J. E., 491
 Kush, J. C., 152, 165
 Kutcher, S., 111
 Kutchins, H., 473
 Kvaal, K., 669
 Kvaal, S., 719
 Kwan, K. K., 262
 Kwilosz, D. M., 696

L

La Rue, A., 180
 Lacey, K., 260

Lachar, D., 47, 64
 Lacks, P., 630, 634–636, 639
 Läge, D., 664
 Lai, Z. C., 218, 225
 Laird, N. M., 108
 LaLone, L., 65, 258
 Lam, C. S., 219
 Lamb, D. G., 273
 Lambert, M. J., 10, 42, 48, 653, 655, 672,
 674
 Landerl, K., 195
 Landrine, H., 61
 Lane, G., 704
 Lang, P. J., 132
 Langbehn, D. R., 648–650
 Lange, A. J., 366
 Lange, R. T., 201, 202
 Langley, J., 81
 Lanyon, R. I., 25
 Lapouse, R., 79
 Larkin, K. T., 135, 136
 Larrabee, G. J., 197, 241
 Larsen, K. H., 132
 Larsen, R. M., 6, 7
 Larson, E. B., 649
 Larzelere, R. E., 110
 Laure, J., 649
 Laurent, H., 649
 Lavelle, L. A., 136
 Lavori, P. W., 105
 Lawrence, S. B., 92
 Lazarus, A. A., 80, 116, 120, 133, 676
 Le Goff, M., 669
 Leahy, B. J., 219
 Leary, T., 90
 Lebow, B. S., 25
 Ledbetter, M. F., 197, 217
 Lee, G. P., 217
 Lee, J., 107
 Lees-Haley, P. R., 277, 428–429
 Lefebvre, M. F., 132

- LeFever, D., 558
 Leffard, S. A., 185
 Lefkowitz, J., 51
 Lehman, A., 649
 Lehne, G. K., 492
 Lei, P., 169
 Leichsenring, F., 674, 675
 Leigh, J. B., 10
 Leighton, A. H., 108
 Lemskey, C., 43
 Lentz, R. L., 684
 Lenzenweger, M. F., 424
 Leon, A. C., 106
 Leong, F., 377
 Lepach, A. C., 231
 Lerner, H. D., 573
 Lessenger, L. H., 261
 Leung, K., 377, 495
 Leura, A. V., 524, 528
 Levak, R. W., 263, 266, 272, 342, 730
 Levant, R. G., 10
 Levenson, M. R., 659
 Levin, H. S., 624
 Levine, D., 40, 50
 Levine, H. M., 695
 Levitt, E. E., 209, 334
 Levitt, J., 654
 Lewak, R., 57, 58, 263, 728
 Lewinsohn, P. M., 107
 Lezak, M. D., 8, 68, 139, 152, 153, 169, 174, 176, 180, 187, 188, 192, 194, 196–199, 215, 625, 626, 630, 631, 645, 708
 Li, Y., 154
 Liang, K., 106
 Liao, Q., 55–56, 82
 Lichtenberger, E. O., 9, 65, 71, 73, 150, 152, 153, 166, 168, 173, 174, 191, 192, 197–199, 204–206, 218, 225, 273
 Lightfoot, S. L., 662
 Lilienfeld, S. O., 6, 7, 517, 518, 520–521
 Lillesand, D. V., 133
 Lilly, R. S., 254
 Lim, C., 702
 Lim, J., 262
 Lin, E. J., 377
 Lindberg, S., 195
 Lindsay, D. S., 130
 Lindsay, K. A., 431
 Ling, W., 683–684
 Linger, M. L., 153, 159
 Lingo, J. C., 295, 334
 Linkersdörfer, J., 195
 Linscott, J., 117
 Lipman, R. S., 655–656
 Litt, M. D., 94, 108–109, 694
 Little, K. B., 254
 Liu, W., 655
 Livesley, W. J., 491
 Lloyd, A. M., 431
 LoBello, S. G., 153
 Lodge, J., 114, 134
 Loe, S. A., 153, 159
 Loenberger, L. T., 43
 Loftus, E. F., 28, 81, 130
 Long, C. J., 152, 619
 Long, J. D., 657
 Lonnemann, J., 195
 Looper, K., 687
 Lopata, C., 636
 Lopez, I. R., 65, 248–249
 Lopez, M. N., 437
 Lopez, S. J., 161, 212
 Loranger, A. W., 92, 427–428
 Lorber, M. F., 136
 Loring, D. W., 219
 Lorr, M., 425
 Love, K. G., 83
 Lovibund, P. F., 669
 Lovibund, S. H., 669
 Lovitt, R., 58
 Lowery, N., 231

Lubin, B., 6, 7
 Luborsky, L., 674
 Lucas, C. P., 109, 110, 681
 Luciana, M., 73
 Luckasson, R. A., 207
 Ludewig, D., 104
 Ludolph, P., 573
 Lumley, V., 121
 Luna-Jonew, L., 256
 Lunnen, K., 655
 Luria, A. R., 8, 153, 617, 618
 Lushene, R., 666, 685
 Lydiard, R. B., 104
 Lyle, O. E., 640
 Lynn, R., 154
 Lyons, J. S., 678

M

Maack, D., 135
 Machamer, J., 796
 Macias, L. G., 659
 MacLane, C., 135
 Macmann, G. M., 152
 Maddox, T., 10, 68
 Madeddu, F., 659
 Maerlander, A., 146
 Magnavita, J. J., 436
 Maher, M. C., 205
 Makatura, T. J., 219
 Maldonado, R., 61
 Malik, M., 102, 672, 676, 681, 741
 Malik, M. L., 704
 Malina, A. C., 223
 Mallinkrodt, B., 695
 Malloy, P. F., 633
 Maloney, M. P., 3, 24, 26
 Mangis, H. A., 185
 Manly, J., 64, 227
 Manos, N., 659
 Manuzza, S., 105
 Marazziti, D., 628

Marchese, M. C., 83
 Marczyk, G. G., 44
 Marek, R. J., 255
 Margo, P. A., 106
 Margolin, G., 126
 Marinkovic, J., 658
 Markon, K. E., 489, 491
 Marks, I. M., 153, 679, 742–743
 Marks, P. A., 250
 Marks, W., 153
 Marley, M. L., 636
 Marlowe, D., 428
 Marmor, D., 197, 219
 Marsh, B. J., 667
 Marshall, M. B., 257, 347
 Marshall, T., 42
 Martin, A., 620
 Martin, M. A., 277
 Martin, P., 111
 Martin, R. C., 217
 Martin, T. A., 490, 491
 Martin-Cannici, C., 383
 Martinez, S., 659
 Martinez-Aran, A., 106
 Maruish, M. E., 34, 42, 68, 250, 272, 346, 653, 654, 677
 Mascolo, J. T., 204
 Maser, J. D., 116
 Mash, E. J., 68
 Masling, J. M., 520
 Masling, M., 523–524, 559
 Mason, B., 520
 Massons, J.M.D., 111
 Masur, J. D., 106
 Matarazzo, J. D., 3, 6, 7, 78, 200, 211
 Mather, N., 47
 Matheson, D. W., 208
 Mathur, S. R., 120
 Matson, M., 231
 Mauger, P. A., 251
 Maugherman, A. S., 25

- Maurer, S. D., 83
 May, D. J., 104
 May, M. A., 115
 Mayes, S. D., 186, 205
 Maziade, M., 104
 McBlaine, D. D., 261
 McBurnett, K., 177–178
 McCabe, D. P., 188
 McCall, R. B., 156
 McCallum, R. S., 64–66, 155
 McCann, J. T., 425, 428
 McCarthy, R. H., 18, 133
 McClelland, D. C., 42
 McClelland, G. M., 678
 McClelland, J., 620
 McCloskey, G., 146
 McConnelloughy, E. A., 700
 McCormick, I. A., 133
 McCown, W., 551, 569
 McCoy, E.L.B., 185
 McCrady, B., 684
 McCrae, R. R., 252, 489–496
 McCrea, M., 652
 McCully, E., 254
 McDaniel, M. A., 83, 702
 McDermott, P. A., 152, 165, 169
 McDougall, A., 551
 McDowell, C. J., 518, 519
 McDowell, I., 94
 McEvoy, L. F., 108
 McFall, R. M., 133
 McFie, J., 198, 199
 McGill, K., 373–374
 McGlyn, F. D., 132
 McGrath, J. J., 116–117
 McGrath, R. E., 31, 256, 257, 334
 McGregor, P., 114
 McGrew, K. S., 47, 627
 McHugh, P. R., 94, 108, 109, 624
 McIntosh, J. A., 634
 McKay, C., 648
 McKay, J. R., 104, 649
 McKenna, T., 340
 McKenna-Foley, J. M., 657
 McKinley, J. C., 243, 246, 247–248, 338, 492
 McLaughlin, S. C., 129, 131, 132
 McLean, J. E., 71, 140, 150–151
 McLean, P., 687
 McLeod, C. C., 41–42
 McMinn, M. R., 58, 72–74, 273
 McNary, S., 649
 McNulty, J. L., 19, 250, 253, 254, 260, 265, 268, 340
 McShane, D. A., 156
 McSwan, K. L., 649
 Meador, K. J., 217
 Meagher, M. W., 663
 Meagher, R. B., 7, 67
 Meagher, S., 42, 432
 Medlin, W. J., 634
 Meehl, P. E., 30
 Megargee, E. I., 65, 72
 Mehrabian, A., 88
 Meijer, R. R., 658, 664
 Meillon, C., 669
 Melzack, R., 67
 Mendelsohn, G. A., 45
 Mendelsohn, S., 107
 Mendelson, M., 662, 663
 Mendoza, J., 719
 Menges, D., 201, 684
 Merchant, A., 108, 109
 Meredith, K., 693, 694, 697, 698
 Merette, C., 104
 Merluzzi, T. V., 114, 117, 132
 Merrell, J., 255
 Merry, W., 693, 694, 697, 698
 Mervielde, I., 493
 Merwin, M., 217
 Messick, S., 18, 20, 23
 Metraux, R. W., 546

- Metz, C., 107
- Meyer, A., 78, 92
- Meyer, G., 24
- Meyer, G. J., 3, 25, 257, 273, 355, 514, 516–521, 523, 524, 527, 541, 547, 559, 579, 581–584, 587, 593, 602, 604, 605, 612
- Meyers, J. E., 618, 629
- Meyers, K. R., 618, 629
- Meyers, L. S., 254
- Michaels, M., 719
- Michelson, A., 698
- Micucci, J. A., 260
- Middlestadt, S. E., 82
- Mihura, J. L., 51, 514, 517–519, 521, 523–527, 541, 547, 549–551, 553–555, 558–565, 567–571, 573–577, 579–581, 593
- Milberg, W., 198
- Milberg, W. P., 141, 152–153, 739
- Milden, R. S., 573
- Miles, C. C., 247
- Miller, A. S., 551
- Miller, D. N., 7
- Miller, H. L., 260, 519
- Miller, H. R., 427
- Miller, J., 279
- Miller, J. A., 185
- Miller, J. B., 225–226
- Miller, K. B., 340
- Miller, L., 198
- Miller, L. S., 257
- Miller, P., 44
- Miller, S. W., 227, 617
- Miller, T., 493–494
- Miller, T. L., 674
- Miller, W. R., 700
- Millis, S. R., 219, 223, 238
- Millon, C., 19, 67, 417, 418, 421
- Millon, T., 6, 7, 19, 42, 47, 67, 417–419, 421–426, 429, 430, 432, 434, 436, 476, 689
- Milne, C. R., 206, 698
- Milne, D., 116
- Milne, L., 206
- Miltnerberger, R. G., 121
- Minassian, A., 551
- Miranda, J., 702
- Mischel, W., 115
- Mitrushina, M., 617
- Mittenberg, W., 173, 177, 199, 202, 241, 616
- Mizes, J. S., 132
- Mkhize, N. J., 659
- Moberg, P. J., 618
- Mock, J., 662, 663
- Moes, E. J., 58, 726, 742
- Moffitt, T. E., 81
- Mogel, S., 212
- Mohatt, G. V., 702
- Mohr, D., 693, 694, 697, 698
- Mohr, D. C., 686
- Mohs, R. C., 620
- Mojtabai, R., 665
- Mold, J. W., 648, 650
- Moliero, C., 81, 672, 676, 698
- Moll, K., 195
- Mollner, N. R., 205
- Monachesi, E. D., 251
- Monahan, J., 45
- Monk, M. A., 79
- Monnette, G., 201
- Monson, R. R., 108
- Moon, G. W., 159, 160
- Mooney, K., 118
- Moore, T. M., 664
- Moos, R. H., 493, 695
- Moreland, K. L., 3, 55
- Morey, L. C., 257, 371, 373–375, 377, 379, 382, 388, 391, 415, 492, 677
- Morgan, C. D., 437
- Morgan, G., 18
- Morlan, K. K., 659

- Morris, J. C., 204, 620
 Morris, R., 141, 146, 171
 Morris, R. L., 668
 Morrison, T. L., 256
 Morrow, L. A., 188, 620
 Mortenson, E. L., 659
 Moscoso, M. S., 669
 Moser, D. J., 648–650
 Moser, J. S., 377
 Moser, R. S., 649
 Mossman, D., 658
 Most, R. B., 55
 Mostofsky, D., 42
 Motowidlo, S. J., 83–84
 Mouton, A., 425
 Muchinsky, P. M., 83
 Muderrisoglu, S., 428
 Mueller, H. H., 208
 Mueller, T. I., 106
 Mulvey, E., 45
 Mundia, L., 262
 Munley, P. H., 251
 Munoz, R., 101
 Murillo, L. G., 548, 559
 Murphy, J. M., 108
 Murphy, L. L., 10, 68
 Murphy, M. J., 10
 Murphy, S. A., 657
 Murray, R. A., 153
 Murray, S. C., 549
- N**
- Nadolne, M. J., 619
 Nagaya, T., 665
 Naglieri, J. A., 155
 Nagy, J., 494
 Nahulu, L. B., 669
 Nakamura, B. J., 683
 Näring, G., 665
 Narow, W. E., 109
 Nathan, J. S., 3, 6, 7, 73, 139, 215, 243, 417, 513, 615, 663, 675, 683
 Nathan, P. E., 683, 975
 Neath, I., 188
 Neff, D. F., 687
 Neher, L. A., 52
 Neighbors, H. W., 82
 Neimeyer, R. A., 696
 Neisser, U., 21, 64, 139
 Nel, A., 665
 Nelson, C., 25
 Nelson, D. V., 669
 Nelson, E. A., 42, 653, 677
 Nelson, H., 201
 Nelson, J. R., 120
 Nelson, N. W., 277
 Nelson, R. E., 116
 Nelson-Gray, R. O., 114, 116, 684
 Nesse, R., 82
 Nestadt, G., 106, 108, 109
 Neufeld, K., 108
 Nevid, J. S., 133
 Newham, J. J., 668
 Newman, A. C., 198–199
 Newman, F. L., 654
 Newport, K., 634, 636
 Newsom, C. R., 513
 Nezu, A. M., 114, 122
 Nezu, C. M., 114, 122
 Nezworski, M. T., 6, 7, 517, 520, 521
 Ngo, L. A., 201
 Nguyen, L., 55–56, 66, 82
 Nichols, D. S., 263, 279, 342, 729
 Nicholson, R., 45
 Nicholson, R. A., 257, 273
 Nickel, E. J., 662
 Nieberding, R., 3, 6, 139, 513
 Niederehe, G., 623, 628
 Nielsen, L., 743
 Nishida, A., 684
 Nishimura, S. T., 669
 Nishimura, Y., 684
 Nissen, M. J., 620

Nitch, S. R., 649
 Noh, S., 659
 Nojima, S., 664
 Noland, R. M., 377
 Norcross, J. C., 581, 672, 674, 675, 679,
 680, 683, 700, 742
 Nordberg, S. S., 653
 Nordhaus, I. H., 669
 Norling, D. C., 135
 Norman, D. K., 438, 439
 Norman, S., 628
 Norris, M. P., 663
 Novy, D. M., 669
 Noyes, R., 104, 428
 Nuevo, R., 665
 Null, C., 252
 Nussbaum, P. D., 43
 Nutzinger, D. O., 657
 Nwosu, A., 665

O

Oas, P., 635
 Ober, B. A., 148, 618
 Oberholzer, E., 415
 O'Brien, W. H., 114, 116–117, 118, 119
 O'Donohue, W., 10, 42
 O'Donohue, W. T., 708
 Oei, T.P.S., 687
 Offord, K. P., 250
 Ogard, E. M., 256
 Ogles, B. M., 653
 Okazaki, Y., 684
 Okiishi, J., 655
 Okun, A., 668
 O'Leary, A. M., 675
 O'Leary, K. D., 118
 Oliver, J. M., 662
 Oliveri, M. E., 687
 Ollendick, T. H., 117, 132, 683
 Olsen, L. R., 659
 Olson, R. E., 431
 O'Neal, J. H., 687

Ones, D. S., 73, 140
 Organista, K. C., 702
 O'Riley, A. A., 665
 Orne, M. T., 118
 Ortiz, S. O., 166
 Ortwein-Swoboda, G., 106
 Osberg, T. M., 25, 67
 O'Shanick, G. J., 257, 521
 Osman, A., 664, 665
 Ost, L. G., 114
 Ostendorf, F., 494
 Osterrieth, P. A., 618, 629
 Othmer, E., 86, 90, 93, 109, 662
 Othmer, S. C., 86, 90, 93
 Otto, M. W., 675
 Otto, R. K., 43–45
 Overall, J. E., 624
 Owen, A. E., 667
 Ownby, R. L., 724–727, 743

P

Paap, M. C., 658
 Pace, T. M., 260
 Pachet, A., 649
 Padesky, C. A., 118
 Paez, P., 107
 Page, H. A., 549
 Pagel, M., 677
 Pancoast, D. L., 250
 Pandleton, M. G., 120
 Pantle, M. L., 551, 572
 Panzarella, C., 695
 Paolo, A. M., 148
 Park, H. I., 520
 Park, N., 741
 Parker, K.C.H., 251, 518, 520, 523
 Parker, W. D., 494
 Parklinson, D., 427, 620
 Parks, C. W., 134
 Parloff, M. B., 687
 Parnell, T., 256

- Parrish, C. L., 491
 Parry, G., 696
 Parsons, M. W., 708
 Parsons, O. A., 201
 Pascal, G. R., 634
 Paski, B., 658
 Pasupathi, M., 493
 Patinella, L., 159
 Patrick, C. J., 136
 Patterson, C. H., 88, 674
 Patton, C., 177, 241
 Patton, D. E., 241, 342, 648, 650
 Patton, W., 342
 Paul, C. A., 173, 177
 Paul, G. L., 673, 684, 703
 Pauls, C. A., 497
 Pauls, F., 231, 232
 Paulson, J., 116
 Pauly, H., 195
 Payeur, R., 104
 Peacock, E., 657
 Peacock, M., 114
 Pearson, N. A., 47, 144, 173, 198, 219,
 221, 222, 226, 227, 234, 240, 241,
 335, 338, 629
 Pechura, C., 42
 Peek, C. J., 10, 42
 Pendleton, M. G., 119–120
 Penick, E. C., 109, 662
 Penley, J. A., 654, 665
 Penley, J. S., 654, 665
 Penney, A. M., 188
 Penney, D. L., 58
 Penrod, S., 28
 Perez, R. G., 111
 Perkins, D. N., 156
 Perry, J. C., 687
 Perry, J. N., 73, 272, 427, 551, 681
 Perry, W., 551
 Persons, J. B., 118, 680
 Perugini, M., 659
 Petermann, F., 154, 231, 657, 659
 Peterson, C., 741
 Peterson, C. A., 431
 Peterson, L. M., 34, 68
 Petoskey, M. D., 7, 47
 Petrocelli, J. V., 700
 Peveler, R. C., 657, 658
 Pfflner, L. J., 177–178
 Pfohl, B., 741
 Phansalkar, S., 619
 Phillips, J., 983
 Phillips, L., 558
 Piasecki, T. M., 82, 117, 123
 Picchetti, M., 628
 Piedmont, R. L., 494–497
 Piers, C. C., 518, 519
 Piersma, H. L., 424, 657
 Pike, P. L., 261
 Pilkonis, P. A., 687
 Pilowski, I., 67
 Pincus, A. L., 493
 Pincus, H. A., 9, 42, 116
 Piotrowski, C., 3, 7, 72, 243, 250, 417,
 515, 663
 Piotrowski, Z. A., 550, 551
 Plancherel, B., 649
 Plas, J. M., 156
 Plutchik, R., 107
 Pogge, D. L., 334
 Pollack, M. H., 675
 Ponsford, J. L., 618, 619, 624, 625
 Ponterotto, J. G., 61, 65, 82
 Pope, K. S., 51, 57, 248, 256, 263, 728
 Porché-Burke, L., 702
 Porter, E. H., 78
 Potvin, O., 669
 Powell, B. J., 109, 662
 Powis, D., 334
 Poythress, N., 45
 Prabucki, K., 107
 Prange, M. E., 110

Presley, G., 527
 Preston, J., 687
 Preti, E., 659
 Pretzer, J. L., 133
 Prevatt, F., 521, 577, 579
 Price, J. R., 428–429
 Price, L., 219, 238
 Prifitera, A., 64, 154, 161, 200, 211, 217
 Prigatano, G. P., 621, 624
 Primoff, E. S., 55
 Prince, J. P., 708, 740
 Prinz, U., 657
 Prochaska, J. O., 676, 680, 699, 700
 Pruitt, D. L., 250
 Pruitt, J. A., 7
 Prunas, A., 659
 Prus, J. S., 743
 Puente, A. E., 3, 6, 7, 73, 139, 215, 243, 417, 513, 615, 663
 Puig-Antich, J., 107
 Purisch, A. D., 8, 617
 Putzke, J. D., 73

Q

Quah, S. L., 702
 Quarington, B., 520
 Quevedo, K. M., 262
 Quigley, B. D., 497
 Quillan, J., 133
 Quilty, L. C., 664

R

Rabin, L. A., 3, 8, 215, 629
 Radar, D. E., 159
 Rana, D. K., 635
 Rand, K. L., 129, 741
 Randolph, C., 645–652
 Ranieri, W. F., 663, 664
 Rankin, L. A., 133
 Rao, S. M., 173, 616
 Rapaport, D., 515, 548
 Rapport, L. J., 188

Ratcliff, K. S., 107
 Rathus, S. A., 119, 133
 Ray, G. E., 153
 Raymond, V., 104
 Razani, J., 617
 Read, J. D., 130
 Realo, A., 494
 Reaume, W. M., 657
 Reddon, E. M., 120, 149
 Reed, G. M., 5, 55, 74
 Rees, A., 675
 Reese, J. B., 527
 Regier, D. A., 109, 427–428
 Reheiser, E. C., 667, 668
 Reich, J. H., 104, 428
 Reich, W., 110, 111
 Reichenberg, L. W., 683
 Reid, D. B., 217
 Reinares, M., 106
 Reinecke, M. A., 152, 186
 Reise, S. P., 664
 Reister, G., 659
 Reitan, R. M., 8, 176, 192, 198, 199, 574, 616, 617, 626, 632
 Reller-Geary, L., 568
 Renjan, V., 700
 Rescoria, L. A., 47
 Resnick, J., 518, 519
 Retzlaff, P., 45, 426, 432, 434–436, 438
 Rey, A., 618, 629
 Reyes, E., 132
 Reynolds, C. R., 47, 64, 66, 140, 154, 201, 205, 634
 Reynolds, M. R., 150
 Reynolds, S., 493, 675
 Rhoades, H. M., 624
 Rhode, R., 107, 525
 Rice, A. S., 109
 Richard, D.D.S., 226
 Richey, C. A., 133
 Richter-Appelt, H., 658

- Rickels, K., 655–656
 Ricker, J. H., 223
 Riddle, M. A., 106
 Ridgeway, E. M., 554
 Ridley, C., 684
 Riemann, R., 491, 496
 Riggs, D. S., 667
 Rinaldo, J., 273
 Riskind, J. H., 104
 Rissmiller, D. J., 662, 665
 Ritschel, L. A., 741
 Robbins, R. R., 260
 Robbins, T. W., 73
 Roberts, B. W., 20
 Roberts, J., 104
 Roberts, L. W., 740–741
 Roberts, M. L., 120
 Roberts, R., 10, 683, 741
 Robertson, G. J., 47, 55, 627
 Robin, R. W., 260
 Robins, D. L., 43
 Robins, E., 101
 Robins, L. N., 8, 85, 107–109
 Robinson, D., 81
 Robinson, J. D., 42
 Robinson, L. H., 273
 Rock, R. C., 256
 Rodell, J. L., 546
 Rodenhauser, P., 93
 Rodgers, D. A., 253
 Rodgers, J. D., 636
 Rodrigue, J. R., 743
 Roeurer, G., 415
 Rogers, C. R., 78, 674
 Rogers, P. A., 669
 Rogers, R., 8, 45, 86, 93, 94, 102, 104,
 105, 107, 111, 273, 277, 373–375,
 377, 382, 383, 426, 437
 Rohde, P., 107
 Rohlunan, D. S., 73
 Rohwer, W. D., 618, 625
 Roid, G. H., 154, 217
 Rolland, J., 494
 Rollman, B. L., 10, 42
 Romanczyk, R. G., 118
 Romanoski, A. J., 108, 109
 Rorschach, H., 514, 515
 Rose, M. P., 132
 Rosen, B. M., 79
 Rosenberg, R., 42
 Rosenhan, D. L., 741
 Rosenstein, L. D., 620
 Rosenthal, R., 51, 52, 54, 62, 85, 118
 Rosqvist, J., 135
 Ross, L. D., 83
 Rosseel, Y., 664
 Rossell, S. L., 649
 Rossi, G., 425
 Rossi, J. S., 700
 Rossini, E. D., 635
 Rössler, W., 662
 Roth, D. L., 141
 Roth, L. H., 45
 Rothman, S., 139, 141
 Rothschild-Yakar, L., 527
 Rotzien, A. L., 45
 Rouaud, O., 669
 Rouse, S. V., 340
 Roussi, P., 665
 Rowe, E. W., 7, 47, 634
 Rowzee, R. D., 669
 Roy, A. A., 104
 Roy, E. A., 218–219
 Rózsa, S., 659
 Rude, S. R., 675
 Ruegg, R. G., 94
 Ruscio, J., 29
 Rush, A. J., 662, 663
 Rush, J. D., 30, 32, 72
 Rushton, J. P., 62, 376
 Russ, S. W., 50
 Russell, E. W., 8, 199, 216, 618

- Rutherford, M. J., 104
Rutherford, R. B., 120
Rutter, M., 79
Ryan, C., 620, 695
Ryan, C. M., 104, 188
Ryan, J. J., 148, 150–151, 173, 177, 211, 212, 225, 226
Ryan, M. K., 57, 161
Ryder, A. G., 496
Ryder, K. A., 649
Rzasa, S. R., 124, 127
- S**
Saal, W., 665
Saccuzzo, D. P., 10, 119, 154
Sackett, P. R., 64, 66
Sacks, N. R., 105
Sadock, B. J., 93, 708
Sadock, V. A., 93, 708
Sahakian, B. J., 73
Saklofske, D. H., 64, 161, 202
Salekin, R. T., 256, 375
Sales, J., 44
Salovey, P., 81
Sam, H. F., 702
Sameroff, A. J., 156
Sammons, M. T., 675, 685
Samuels, J. F., 106
Sanders, R. D., 658
Sanderson, C., 493
Sandoval, J., 154
Sandowicz, M., 71, 72, 698
Sandy, L. R., 727
Sanglade, A. A., 565, 566, 568, 580
Sanislow, C. A., 497
Sankis, L. M., 431
Santrock, J. W., 679, 742
Saperia, D., 625
Sapienza, M. M., 58, 742
Saran, M., 619
Sarason, B. R., 695
Sarason, I. G., 695
Sarno, I., 659
Sartorius, N., 81, 427–428
Saslow, G., 120, 121
Sattler, J. M., 46, 51, 52, 89, 154, 160, 161, 173, 174, 187, 204, 209, 211, 634, 638, 641, 643, 708, 719
Satz, P., 212
Savage, C. R., 134
Savitz, K. L., 657, 658, 660
Sawalani, G., 129
Sayers, S. L., 121, 123, 133
Saylor, C. R., 634
Sbordone, R. J., 152, 619, 620, 630
Scandell, D. J., 496, 497
Scepansky, J. A., 493
Schafer, J., 548
Schaffer, T. W., 355, 518, 523, 527, 541
Schaible, D., 551
Schalock, R. L., 207
Schatz, P., 649
Schechter, M. D., 623–624
Scheltens, P., 203
Schenk, P. W., 255
Scheupback, C. S., 620
Schinka, J. A., 65, 202, 258, 261, 374, 496, 497, 620, 623
Schirling, J., 663
Schleifer, S. J., 110
Schmidt, F. L., 64, 83, 139, 140, 494, 686
Schmidt, H. O., 558
Schmidt, J., 492–493
Schmidt, M., 217
Schmidt, N. B., 675
Schmitz, N., 657–659
Schnakenberg-Ott, S. D., 226
Schneider, L. M., 141
Schoenbaum, M., 10, 42, 200
Schoenberg, M. A., 200
Schoenberg, M. R., 201, 202, 438, 648, 650
Schoevers, R. A., 683–684
Schrader, W. B., 64

- Schrank, F. A., 47, 627
 Schreiber, M., 551
 Schuerger, J. M., 156
 Schulman, I., 549
 Schulte-Körne, G., 796
 Schultz, C. B., 51
 Schultz, H., 657
 Schumaker, J. F., 42, 73, 74, 81, 273, 718
 Schutte, C., 225–226
 Schwab, J. J., 668
 Schwab-Stone, M. E., 109
 Schwartz, G. F., 42, 251
 Schwartz, L., 210
 Schwarz, S. P., 25, 254, 687
 Scott, J. G., 648–650
 Scott, R. L., 253–254
 Seelen, J., 248
 Seeley, J. R., 107
 Seeman, W., 250
 Seever, M., 6, 7
 Segal, D. L., 665, 668, 669
 Seifer, R., 156
 Selagea, F., 679, 742
 Seligman, L., 132, 683
 Seligman, M.E.P., 674, 741
 Sellbom, M., 254–256, 342
 Semmel, A., 132
 Sewell, K. W., 45, 107, 277, 373–374, 377, 382, 383
 Sewitch, T., 132
 Shaevel, B., 259
 Shaffer, D. M., 109, 110
 Shaffer, T. W., 516, 517, 523
 Shalev, A. Y., 656
 Shalit, B., 550
 Shanahan, S. F., 659
 Shanfield, S., 677
 Shanley, L. A., 431
 Shapiro, D. A., 674, 675, 696
 Shapiro, P. N., 28
 Shapiro, S., 108, 109
 Sharp, C., 163, 164
 Shatin, L., 559
 Shaver, P. R., 493
 Shaw, B. F., 662
 Shea, M. T., 687, 695
 Shea, T., 687
 Shear, M. K., 104
 Shedler, J., 428
 Sheedy, D., 649
 Sheehan, E. P., 436
 Sheer, K. J., 108
 Sheikh, A. A., 134
 Sheldon, T., 678
 Shen, L., 696
 Shen, Y., 659
 Shepard, P. H., 255
 Sherbenou, R. J., 155
 Sherman, E.M.S., 68, 219, 227, 623, 628–629
 Sherman, R. H., 51
 Sherry, A. R., 262
 Sheslow, D., 629
 Shimokawa, K., 653
 Shimura, M., 659
 Shinfuku, N., 659
 Shkalim, E., 262
 Shneidman, E. S., 254
 Shogren, K. A., 207
 Shondrick, D. D., 261
 Shores, E. A., 649
 Shorkey, C. L., 132
 Short, R. H., 208
 Shrauger, J. S., 25, 67
 Shultz, J., 492–493
 Siddiqui, N., 373, 377
 Sieber, K. O., 254
 Siefert, C. J., 373–374
 Siegel, A. W., 623–624
 Siegel, L., 263, 730
 Siegle, G., 132
 Siegler, I. C., 492

- Sigelman, B. E., 758, 759
 Silva, P. A., 81
 Silver, D., 107
 Silver, E., 45
 Silver, E. J., 668
 Silver, M. J., 51
 Silver, S. E., 110
 Silverstein, A. B., 209
 Sim, K., 702
 Simões, A., 494
 Simon, L., 659
 Simon, W. E., 52, 740–741
 Simons, A. D., 342, 687
 Simons, R., 342
 Simonsen, E., 425
 Sinclair, S. J., 373–374
 Singer, B., 674
 Singer, J. A., 493, 565
 Singer, P., 90
 Singer, R. M., 620
 Sisson, R. A., 624
 Skodol, A. E., 497
 Skrandal, A., 657, 658
 Slate, R. J., 153, 159, 160
 Sledge, W. H., 104
 Sletten, I. W., 94, RR
 Slick, D. J., 623, 628–629
 Smart, D. W., 653
 Smerbeck, A. M., 636
 Smith, B., 519
 Smith, C., 260, 527
 Smith, D., 7
 Smith, G. L., 110
 Smith, G. T., 6, 18, 23
 Smith, J. Z., 702
 Smith, K. G., 669
 Smith, M., 132
 Smith, M. C., 7
 Smith, M. L., 674
 Smith, R. L., 251
 Smith, S. R., 330, 579
 Smith, T. B., 677, 680
 Smith, T. P., 679, 702, 742
 Smithberger, A., 57
 Smith-Wilson, R., 659
 Snitz, B. E., 25
 Snow, R. E., 51
 Snow, W. G., 200
 Snyder, C. R., 741
 Snyder, L. E., 204
 Snyder, P. J., 43
 Snyder, W. V., 78
 Snyderman, M., 139, 141
 Sobel, D. S., 42
 Sobol, A. M., 108
 Sofia, C., 649
 Sokolov, R. L., 152
 Solhan, M., 81, 117
 Sommer, J. F., 683–684
 Sommer, R., 679, 742
 Sommers-Flanagan, J., 89, 90, 93
 Sommers-Flanagan, R., 89, 90, 93
 Song, W., 377, 495
 Song, X., 698
 Soref, E., 57, 273
 Sotsky, S. M., 687
 Spanhel, C., 759
 Sparks, J. A., 653
 Sparrow, S. S., 47
 Spellacy, F., 623, 628–629
 Spence, N., 67
 Spengler, P. M., 25
 Spiegler, M., 11, 114, 121, 130
 Spielberg, C. D., 72, 666–669, 685
 Spies, R. A., 10, 68
 Spiker, D. G., 106
 Spiro, A., 659
 Spisto, M. A., 205
 Spitzer, R. L., 8, 81, 101–105, 257, 428, 741
 Spitznagel, E. L., 108
 Spreen, O., 68, 227

- Springer, S. P., 210
 Sprock, J., 551
 Srisayekti, W., 665
 Stafford, K. P., 261
 Stanley, M. A., 668
 Stanton, H. C., 205
 Starace, N., 683
 Starcevic, V., 658
 Startup, M., 675
 Steadman, H. J., 45
 Steel, P., 492–493
 Steen, T., 741
 Steer, R. A., 67, 104, 652, 662–665
 Steffen, A. M., 674, 675
 Stein, J. A., 695
 Stein, L.A.R., 206, 340
 Stein, M. A., 152
 Stein, R. K., 668
 Steinberg, D. L., 114
 Steinberg, M., 102, 103
 Steiner, J. L., 104
 Stejskal, W. J., 7, 520
 Steketee, G., 675
 Stellman, J. M., 683–684
 Stellman, S. D., 683–684
 Stern, R. A., 616
 Stern, S. L., 106
 Sternberg, R. J., 66, 139, 141, 152, 153, 156
 Sternklar, S., 559
 Stiers, W., 796
 Stillerman, L., 659
 Stolzman, R. K., 108
 Stompe, T., 106
 Stone, G. E., 657
 Stone, S. V., 493
 Storandt, M., 636
 Storch, E. A., 743
 Storyback, D., 73
 Stout, C. E., 653
 Stout, J., 633
 Stover, E., 620
 Strack, S., 422, 425
 Strauss, E., 68, 219, 227, 241, 617, 623, 628–629, 648
 Strauss, J. S., 684
 Strauss, M. E., 493
 Stredny, J. K., 7, 44, 139, 243, 417, 521, 579
 Stredny, R. V., 521
 Streiner, D. L., 427
 Stricker, G., 44
 Stringer, A. Y., 619
 Strobl, R., 106
 Stroop, H. H., 618, 625
 Strupp, H. S., 78, 677
 Stuart, H. J., 257, 521
 Stukenberg, K. W., 658, 663
 Sturmey, P., 708
 Stutts, M. L., 198–199
 Subica, A. M., 663, 664
 Suchy, Y., 618
 Sue, D. W., 702
 Suen, H. K., 124, 127
 Sullivan, D. S., 51
 Sullivan, E. V., 620
 Sullivan, T. E., 217
 Summerfeldt, L. J., 85
 Sumpter, J., 273
 Sun, H., 495
 Sundberg, N. D., 718
 Sundsmo, A., 135
 Supernant, A. M., 188
 Suprenant, A. M., 188
 Suttell, B. J., 634
 Suzuki, L. A., 61, 65
 Svetlana, A. B., 657
 Swart, E. C., 109, 335
 Swartz, H. A., 104
 Swartz, M. S., 109
 Sweet, J. J., 277, 618, 620
 Swift, J. K., 702

- Swindle, R., 684
 Swinson, R. P., 668
 Sydeman, S. J., 667, 668
 Szasz, T., 54
 Szekely, C., 687
 Szirmák, Z., 494
- T**
- Tabrizi, M. A., 107
 Taft, R., 28
 Takahashi, H., 665
 Talaga, M. C., 687
 Talebi, H., 672, 681, 698
 Tallent, N., 719, 723, 724, 726
 Talley, P. F., 677
 Tamkin, A. S., 549
 Tan, S. Y., 659
 Tanii, H., 684
 Tanner, C., 623
 Tarescavage, A. M., 255, 256
 Tarrier, N., 681
 Tasca, G. A., 373–374
 Tashman, N. A., 114
 Tasman, A., 683
 Taub, G. E., 150
 Taub, J., 10, 654, 677
 Taylor, J., 51
 Taylor, M. J., 64, 227, 617
 Taylor, N., 519
 Taylor, R. J., 82
 Taylor, S., 687
 Teal, M., 152, 180
 Tebes, J. K., 104
 Tehan, G., 188
 Teitelman, K. B., 665
 Tellegen, A., 19, 243, 249, 250, 252,
 254–256, 265, 303, 342, 347, 353,
 355, 366, 697
 Teng, E. L., 625
 Terman, L. M., 247
 Terracciano, A., 489, 491
 Terrell, F., 51
 Terrell, S. L., 51
 Test, M. A., 657
 Thakler, C. D., 620
 Thase, M. E., 687
 Thelen, M. H., 7, 132
 Thomann, J., 676
 Thomas, E. A., 520, 554
 Thomeer, M. L., 649
 Thompson, L., 676
 Thornton, D., 45
 Thorp, R. G., 118
 Tibon Czopp, S., 527
 Tilburg, W. V., 683–684
 Tillbrook, C. E., 107
 Tilley, B. C., 649
 Timbrook, R. E., 45, 64, 256, 260
 Timko, C., 493
 Toepfer, S., 259
 Tokudome, S., 665
 Tolbert, H., 259
 Tolman, A. O., 45
 Tombaugh, T. N., 94
 Tomcho, T. J., 121, 123
 Toner, B. B., 436
 Tonigan J. S., 700
 Tonsager, M. E., 58, 728, 729
 Toomey, J. A., 636
 Torgerson, S., 104
 Toruno, J. A., 262
 Tranel, D., 43, 59, 139, 215, 626, 708
 Treadwell, K., 129, 130, 132
 Tremont, G., 202
 Tress, W., 657, 659
 Trimble, J. E., 702
 Tripp, G., 114, 134
 Tröster, A. I., 628
 Truax, C. B., 78
 Trueblood, W., 217
 Trull, T. J., 81, 117, 375, 428, 429
 Trumbetta, S. L., 259
 Trumbo, D., 82, 83

Tsai, D. C., 261
 Tsushima, V. G., 261
 Tsushima, W. T., 261
 Tucker, R. K., 120
 Tulskey, D. S., 197, 219
 Turbanisch, U., 659
 Turk, D., 81
 Turkat, I. D., 427
 Turner, S. M., 5, 55–57, 74, 119, 687
 Twentyman, C. T., 133
 Tyrano, S., 107
 Tziner, A., 82

U

Uhlenhuth, E. H., 655–656
 Ullman, L. P., 115
 Ulrich, L.P., 82, 83
 Ulstad, K. L., 98, 377
 Ulstein, I., 669
 Umphress, V., 655
 Underhill, A. T., 153
 Unoka, Z., 658, 659
 Urbán, R., 658
 Urbina, S., 21, 64, 139
 Urgelles, J., 431–432
 Ustad, K., 373–374, 377, 382

V

Vagg, P. R., 666, 685
 Vaglum, P., 659
 Vakil, E., 658
 van Belle, G., 620
 van den Berken, J.H.L., 29
 van den Bosch, R. J., 679
 Van Denburg, E., 425, 431, 432, 689
 van der Veld, W. M., 665
 Van Hasselt, V. B., 668, 669
 Van Zomeren, A. H., 625
 Vanderploeg, R. D., 202
 Vandiver, T., 108
 VanDyke, S., 225–226
 Vane, J. R., 31

Vanheule, S., 664
 Vásquez, M., 702
 Vassend, O., 657, 658
 Vázquez-Barquero, J. L., 665
 Veiel, H.O.F., 201, 202
 Velicer, W., 700
 Ven Denburg, E., 429
 Verdirame, D., 694
 Vernon, P. E., 28
 Veron-Guidry, S., 116
 Verschell, M. S., 518, 519
 Vieta, E., 106
 Viglione, D. J., 514, 518, 519, 527, 565
 Vigneau, F., 669
 Villacis, L., 665
 Villar-Hoz, E., 659
 Viñas-Poch, F., 659
 Vincent, K. R., 420, 521, 561
 Visser, R.S.H., 618
 Vitacco, M., 277
 Vogt, A. T., 196
 Volker, M. A., 636
 Vollmer, B. M., 702
 von Baeyer, C., 132
 von Hofacker, N., 796
 Vujnovic, R. K., 636

W

Wagner, R. K., 78, 82, 139
 Wagstaff, N., 704
 Waldmann, H. C., 154
 Waldo, M., 659
 Walker, E. A., 42
 Walker, E. J., 558–559
 Walker, M. L., 104
 Walker, R. N., 546
 Wallbrown, F., 696, 699
 Waller, N. G., 494
 Walsh, K., 627
 Walsh, W. B., 256
 Walton, R. G., 677

- Wampold, B. E., 672, 674, 676
Wang, X., 659
Ward, C. H., 377
Ward, L. C., 150–151, 182, 211, 212, 256, 664, 665
Ward, M. P., 3, 24, 26
Ward, T. J., 205
Warren, J. I., 695
Wasserman, J., 64–65
Wasyliw, O. E., 525
Watkins, C. E., 3, 7, 114, 139, 513
Watkins, J. T., 687
Watkins, M. W., 6, 151, 152, 165, 169
Watson, C., 132
Watson, D., 489, 491
Watson, J. C., 672
Watts, D., 273
Watzlawick, P., 79
Way, B. B., 678
Weaver, R. L., 120
Webb, J. F., 263
Webb, W. W., 257
Webster, C. D., 45
Webster, J. S., 188
Wechsler, D., 142–150, 155, 157, 163, 164, 173, 181, 186–188, 191, 192, 194, 196–199, 201, 203, 205, 206, 209, 216–218, 223–225, 227, 526, 629
Wedding, D., 29
Weed, L. L., 116
Weed, N. C., 259, 272, 340
Weinberg, R. A., 139
Weiner, D. A., 678
Weiner, I. B., 25, 513, 514, 516, 519, 520, 522, 541, 545, 554, 555, 557, 561, 565–569, 580
Weingardt, K. R., 742
Weiss, L. G., 64, 150, 154, 219, 227, 238
Weisskopf, M. G., 796
Weissman, A., 131
Weissman, H. N., 256
Weisz, J. R., 110, 683
Weller, R. A., 109
Welner, A., 8
Welsh, G. S., 252
Wenger, A., 426
Werth, T. R., 161, 212
Wertheimer, J. C., 648
Wescott, M. C., 658
Westen, D., 428
Westwood, M., 668
Wetter, M. W., 273, 279
Wetzler, S., 428
Whetzel, D. L., 83
Whisman, M. A., 665
Whitbeck, L. B., 702
White, G. W., 743
White, M. J., 25
White, S., 82
White, T., 616
Whiteford, N. T., 665
Whitehouse, W. G., 114, 695
Whitelaw, S., 700
Whiteman, V. L., 132
Whitney, K. A., 255
Whitworth, R. H., 261
Wickramasekera, I. E., 676
Widiger, T. A., 9, 102, 116, 420, 428, 429, 431
Wiebe, J. S., 654, 665
Wiedel, T. C., 25, 254
Wiederholt, J. L., 155
Wielkiewicz, R. M., 185, 187, 199, 206
Wiener-Levy, D., 568
Wiens, A. N., 80
Wiersma, D., 679
Wiesner, W., 83
Wiggins, J. S., 328–329, 347, 493
Wilcox, N. S., 700
Wilde, M. C., 649
Wilde, N. J., 219
Wilke, C. M., 649, 650

- Wilkinson, G. S., 47, 627
 Willburger, E., 195
 Williams, C. L., 19, 248, 257, 259, 272, 303, 334
 Williams, D. R., 82
 Williams, J., 428
 Williams, J. M., 29, 629
 Williams, J.B.W., 8, 102–104
 Williams, K., 132, 201
 Williams, M. A., 273
 Williams, O. B., 655, 703, 741
 Williams, R. B., 252, 493
 Williams, R. B., Jr., 492
 Williams, R. E., 704
 Williams, R. N., 649
 Williams, T. A., 80
 Williams, W. M., 64
 Williamson, D. A., 116
 Wilson, B. A., 629, 633
 Wilson, R. S., 623
 Wilson, T. A., 80
 Windover, A., 255
 Winfield, I., 109
 Wing, J. K., 81
 Winget, B. M., 52
 Winokur, G., 101, 106
 Wise, S. L., 698
 Witmer, J. M., 51
 Witt, A. C., 124, 156
 Witt, J. C., 124
 Wittchen, H. U., 104, 257
 Wittman, C.L.M., 29, 688
 Wittkowski, A., 668
 Woerner, W., 195
 Wojcik, J. V., 377
 Wolber, G. J., 719
 Wolff, C. B., 659
 Wolfson, D., 8, 175–177, 180, 192, 199, 616, 623, 626, 632
 Wolpe, J., 132, 133
 Wonderich, S., 132
 Wong, S.C.P., 741
 Wood, G., 195
 Wood, J. M., 6, 7, 373–374, 517, 518, 520, 521
 Wood, J. W., 520–521
 Woodcock, R. W., 47, 627
 Woodruff, R. A., 101
 Woodward, J. L., 225
 Woodward, M. J., 432
 Woodworth, R. S., 656
 Woody, R. W., 5
 Wool, L., 106, 373–374
 Worden, B., 201, 684
 Worthington, E. L., 702
 Wright, A. J., 24, 34, 35, 68, 541, 681, 682, 719, 720, 724, 726, 729
 Wycoff, K. L., 204
 Wygant, D. B., 254–255, 347
 Wylie, J. R., 525, 573
 Wyndowe, J., 108
 Wysocki, J. J., 620
- X**
- Xie, D., 495
- Y**
- Yamagata, S., 491
 Yamamoto, I., 664
 Yang, C., 649
 Yang, J., 494, 496, 497
 Yang, M., 741
 Yang, P. H., 61
 Yao, S., 494, 497
 Yasui, M., 683–684
 Yates, B., 10, 654, 677
 Yeh, C. J., 105
 Yokoyama, C., 684
 Yoon, K., 494
 Yoshioka, D. T., 82, 117
 Yost, K., 126
 Young, J., 683
 Young, M. S., 496, 497

Z

Zaballero, A. R., 373–376, 377
 Zabora, J. R., 659
 Zachar, P., 153
 Zakris, J., 109
 Zald, D. H., 25
 Zalewski, C., 3, 7, 417
 Zapata, J. A., 257, 521
 Zaylor, C., 10
 Zebb, B. J., 668, 675
 Zedeck, S., 82
 Zeltzer, L., 42
 Zembar, M. J., 623, 628
 Zevalkink, J., 664
 Zhang, H., 109, 659

Zhang, J., 377, 495
 Zhang, K. A., 664
 Zhang, X., 657
 Zhao, C., 659
 Zhu, J., 150–151, 154
 Zilmer, E. A., 198–199
 Zimbardo, P. G., 133
 Zimmerman, M., 741
 Ziskin, J., 6, 44, 73
 Zlotnick, C., 695
 Zolliker, A., 559
 Zonderman, A. B., 252, 492, 493
 Zotter, D. L., 132
 Zuckerman, E. L., 51, 89, 679, 719,
 742

Subject Index

A

- AAMD Adaptive Behavior Scale, 208
- Acculturation Rating Scale for Mexican Americans, 61
- Achenback Child Behavior Checklist, 47
- Acid profile, 205
- Adaptive Behavior Assessment System/ABAS, 208
- Administration on Intellectual and Developmental Disabilities (AIDD), 207–208
- Advanced Clinical Solutions I* (Pearson), 198
- Affect Adjective Checklist, 667
- African American Acculturation Scale, 61
- Agreeableness (A) scale, 502–503
- American Board of Forensic Psychology, 44
- American Educational Research Association, 18
- American Educational Research Association (AERA), 50
- American Journal of Managed Care*, 40
- American Psychological Association, 18, 40, 45, 53, 56, 59, 86, 207, 273, 683, 708
- Anxiety Disorders Interview Schedule (ADIS–IV), 102, 741
- Anxiety Scale Questionnaire, 667, 668
- Anxiety (ANX) symptom dimension, 661
- Arbitrary inference, 131–132
- Army Alpha, 142
- Army Beta, 142
- Army Group Examinations (Picture Arrangement), 142
- Asian Values Scale, 61
- Assessment, ethical practice of: and communicating test results, 57–58; and competent use of assessment instruments, 55–56; and developing professional relationship, 50–52; interpretation and use of test results in, 56–57; and issues related to informed consent, 52–54; and labeling and restriction of freedom, 54–55; and maintenance of test security and release of test data, 58–59
- Assessment, of diverse groups: and cultural competency, 60; and cultural/racial identity, 60–61; diagnostic issues in, 66–67; interpretation guidelines for, 67; language skills for, 59–60; test equivalence and appropriate use of instruments in, 61–66
- Assessment domains, sample grid of, by tests administered, 738 Tab. 15.2
- Assessment instruments relevant for specific domains, 69–71 Tab. 2.2
- Assessment interview, 77; assets and limitations of, 84–85; case history of, 86–93; developments during 1960s, 79; developments during 1970s, 79–80; developments during 1980s, 80–81; developments during 1990s and into millennium, 81–82; developments during 1940s and 1950s, 78; early developments in,

Assessment interview (*Continued*)
 78; and interpreting interview data, 99–100; issues related to reliability and validity in, 82–84; and mental status evaluation, 93–99; and structured interviews, 100–111; tactics, 89–93

Assessment of Children (Sattler, Sattler, and Hoge), 708

Auditory working memory (short-term memory), 168

Automated Assessment Associates, 272

Automatic Thoughts Questionnaire, 131–132

Axis diagnoses, 103, 104

Axis IV Psychosocial and Environmental Problems, 685

Axis V Global Assessment of Functioning, 685

B

Bakker Assertiveness Inventory, 133

Bannatyne Sequencing, 631

Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS), 620

BASIC-ID. *See* Lazarus BASIC-ID

BASIC-ID (Behaviors, Affects, Sensory experiences, Imagery, Cognitions, Interpersonal relationships, and need for Drugs), 676

Beck Depression Inventory (BDI), 7, 111, 425, 657, 658

Beck Depression Inventory-II (BDI-II), 67, 131–132, 654

Beck Depression Inventory–II (BDI–II), 67, 131–132, 654; interpreting, 665–666; overview of, 662–663; reliability and validity in, 663–665; use of, with diverse groups, 665

Beery-Buktenica Developmental Test of Visual Organization–IV, 630

Behavior Assessment System for Children (BASC–3), 47, 150

Behavior Data, 272

Behavioral assessment, 113–115; and recording of ratings, 129; assets and limitations of, 120–121; and behavioral interviewing, 122–124; behavioral observation in, 124–126; cognitive behavioral assessment in, 129–130; and cognitive self-report inventories, 130–134; and event recording, 127–129; history and development of, 115–117; and interval recording, 127; issues related to reliability and validity, 117–120; and narrative recording, 126; and psychophysiological assessment, 135–136; recommended reading for, 136–137; and recording cognitions, 134–135; strategies of, 121–129

Behavioral Assessment (journal), 116, 117

Behavioral Assessment: A Practical Handbook (Bellack and Hersen), 116

Behavioral interview, 122–124; outline of expected accomplishments for, 124

Bender Gestalt. *See* Bender Visual Motor Gestalt Test

Bender Visual Motor Gestalt Test, 7, 98

Bender Visual-Motor Gestalt Test, Second Edition (Bender–2), 615, 624, 629, 630; administration of, 638–640; assets and limitations of, 637–638; history and development of, 633–636; interpretation guidelines for, 641–645; performance of, among selected clinical populations, 637 Tab. 12.3; reliability and validity in, 636–637; scoring in, 640–641

Benton Visual Motor Retention Test, 629, 640, 641

Block design, 171

Blotto (18th Century game), 514

Boston Diagnostic Aphasia Examination, 626

- Brain damage, assessing: general principles of, 196–199
- Brain impairment, interviewing for, 620–624; and examples of behavioral and emotional changes that may indicate pathological processes in brain, 621 Tab. 12.1
- Bricklin Perceptual Scales, 46
- Brief instruments: and selecting brief instruments, 654–655; for treatment planning, monitoring, and outcome assessment, 653–654
- Brief Symptom Inventory (BSI), 425, 655, 685, 691
- Bulimia Test-Revised, 132
- Butcher Treatment Planning Inventory, 655
- C**
- Caldwell Report, 272
- California Neuropsychological Testing for Neurotoxicity Battery, 620
- California Psychological Inventory (CPI), 7, 25, 72
- California Verbal Learning Test-II, 148, 618
- Cambridge Neuropsychological Test Automated Batteries (CANTAB), 73
- Case formulation, 680–683; Common Function Model for, 682; Complex Model for, 682; Developmental Model for, 681; Diathesis-Stress Model for, 681
- Case history: checklist for assessment interview and, 87 Tab. 3.1; general considerations in, 86–89
- CATEGO (Wing, Cooper, & Sartorius), 81
- Cattel-Horn-Carroll theory of intelligence (CHC), 147, 153
- Central nervous system (CNS), 615, 616, 625, 643
- Child Behavior Checklist (CBCL), 107
- Client characteristics, 675
- Clinical assessment: and assessment of diverse groups, 59–67; and clinical judgement, 26–32; computer-assisted, 72–74; and evaluating psychological tests, 10–23; patterns of test usage in, 6–10; phases in, 32–37; recommended reading in, 37; selecting psychological tests for, 67–72; and validity in clinical practice, 23–26
- Clinical assessment, context of: and academic/educational context, 46–48; ethical practice of, 50–59; and general medical setting, 41–43; and legal context, 43–46; and psychiatric setting, 40–41; and psychological clinic, 48–50; and types of referral settings, 39–40
- Clinical assessment, phases in: and data collection, 33–34; and evaluating referral question, 32–33; and hypothesis testing model for interpreting assessment data, 35 Fig. 1.1; and interpreting data, 34–37
- Clinical decision making, 679–680; and coping style, 692–694; and current life circumstances, 696; and diagnosis, 683–634; and functional impairment, 634–687; and other client characteristics, 701–703; and problem complexity and chronicity, 687–690; and resistance, 696–699; and social support, 694–696; and stage of change, 699–701; and subjective distress, 690–692; and understanding problems, 683
- Clinical Handbook of Psychological Disorders* (Barlow), 707–708
- Clinical judgement: accuracy of, 27–30; and clinical *versus* actuarial predication, 30–32; data gathering

- Clinical judgement (*Continued*)
and synthesis in, 27; and
psychological report, 32
- Clinical Neuropsychology: A Handbook
for Assessment* (Parsons and
Hammke), 708
- Clinical personality patterns (Millon):
Scale 1: Apathetic-Asocial-
Schizoid (AASchd), 439–441;
Scale 2A: Shy-Reticent-Avoidant
(SRAvoid), 441–445; Scale 2B:
Dejected-Forlorn-Melancholic
(DFMelan), 445–448; Scale 3:
Deferential-Attached-Dependent
(DADepn), 448–450; Scale 4A:
Sociable-Pleasuring-Histrionic
(SPHistr), 450–454; Scale 4B:
Ebullient-Exuberant-Turbulent
(EETurbu), 454–457; Scale 5:
Confident-Egotistic-Narcissistic
(CENarc), 457–460; Scale 6A:
Aggrandizing-Devious-Antisocial
(ADAntis), 460–464; Scale 6B:
Assertive-Denigrating-Sadistic
(ADSadis), 464–467; Scale 7:
Reliable-Constricted-Compulsive
(RCComp), 467–470; Scale 8A:
Discontented-Resentful-
Negativistic (DRNegat), 470–473;
Scale 8B: Abused-Aggrieved-
Masochistic (AAMasoc),
473–476
- Clinical syndromes (Millon): Scale A:
Generalized Anxiety, 486;
Scale B: Alcohol Use, 487;
Scale D: Persistent Depression,
487; Scale H: Somatic Symptom,
486; Scale N: Bipolar Spectrum,
486–487; Scale R: Posttraumatic
Stress, 487; Scale T: Drug Use,
487
- Clinician, role of, in conducting
assessments, 3–5
- Clinician's Handbook of Adult Behavioral
Assessment* (Hersen), 121
- Clinician's Handbook of Child Behavioral
Assessment* (Hersen), 121
- Clinician's Handbook of Evidence-Based
Practice Guidelines* (O'Donahue &
Fisher), 707–708
- Cognitive behavioral therapy (CBT),
670
- Cognitive Bias Questionnaire, 131–132
- Cognitive Error Questionnaire, 132
- Cognitive proficiency, 168
- Cognitive self-report measures, 131
Tab. 4.1
- CogScreen, 73
- Collaborative/Therapeutic Assessment:
A Casebook and Guide* (Finn,
Fischer, and Handler), 730
- Compendium of Neuropsychological Tests*
(Strauss, Sherman, and Spreen),
68, 617
- Comprehensive Handbook of
Psychological Assessment
(Volume 3): Behavioral
Assessment* Hayne & Heiby), 121
- Comprehensive System: coding,
531–536; content, 536, 556–563;
determinants, form quality, and
organizational activity in,
532–536; and determinants,
548–556; and giving testing
instructions, 528–529; inquiry,
530; interpretation, 539–581; and
introducing respondent to
technique, 528; and location,
543–547; popular responses, 536;
popular responses used in, plus
proportions of each appearing in
samples of non-patient and patient
protocols, 540 Tab. 11.6; ratios,
percentages, and derivations,
563–581; response (association)
phase of, 529–530; scoring and
interpretive domains for, 544–545
Tab. 11.8; and scoring the
structural summary, 536–539;
special scores, 536; and symbols
and criteria for coding content,
537–539 Tab. 11.5; and symbols
and criteria for coding form

- quality, 535 Tab. 11.4; and symbols and criteria for determinant coding, 533–534 Tab. 11.3; and symbols and criteria used for development quality, 532 Tab. 11.2; symbols and descriptions for special scores in, 542 Tab. 11.7; and symbols used for coding location of Rorschach responses, 531 Tab. 11.1
- Comprehensive Test of Nonverbal Intelligence-2 (CTONI), 155
- Conceptual validity, 26
- Concurrent validity, 19
- Conducting Psychological Assessment: A Guide for Practitioners* (Wright), 681
- Confirmatory bias, 28
- Conflict Resolution Inventory, 133
- Connors–3 Parent and Teacher Rating Scales, 47
- Conscientiousness (C) scale, 503–505
- Consortium to Establish a Registry for Alzheimer's disease (CERAD), 204, 620
- Construct validity, 21–23
- Continuous Performance Test–3 (CPT— 3), 625
- Coping Deficit Index (CDI), 579–580
- Criterion contamination, 21
- Criterion validity, 19–21
- D**
- Delayed Memory Index, 649
- Delis-Kaplan Executive Functioning System, 148
- Dementia Rating Scale, 646
- Depression, 131–132
- Depression Index (DEPI), 579
- Depression (DEP) symptom dimension, 661
- Dhat syndrome, 66
- Diagnosis, 683–684
- Diagnostic and Statistical Manual of Mental Disorders (DSM-5;* American Psychiatric Association), 30, 40, 66, 86, 208, 257, 419, 421, 432, 685, 689, 741
- Diagnostic and Statistical Manual of Mental Disorders (DSM-II;* American Psychiatric Association), 9, 101
- Diagnostic and Statistical Manual of Mental Disorders (DSM-III–R;* American Psychiatric Association), 81, 116, 373
- Diagnostic and Statistical Manual of Mental Disorders (DSM-III;* American Psychiatric Association), 81
- Diagnostic and Statistical Manual of Mental Disorders (DSM-IV;* American Psychiatric Association), 86, 114, 400, 421, 436
- Diagnostic and Statistical Manual of Mental Disorders (DSM;* American Psychiatric Association), 257, 417
- Diagnostic Interview for Children and Adolescents (DICA), 85, 110–111
- Diagnostic Interview Schedule (DIS), 8, 85, 107; adult version, 108–109; child version (Diagnostic Interview Schedule for Children; DISC), 109–110
- DIANO III (Spitzer, Endicott, and Cohen), 81
- Dictionary of Behavioral Assessment Techniques* (Hersen & Bellack), 121
- Differential therapeutics, 675
- Digit Symbol-Coding, 629
- Disability, 619
- Disorders of Personality* (Millon), 419, 421, 423, 432, 436
- Dysfunctional Attitudes Scale, 131–132

E

- Eating Attitudes Test, 132
- Eating Disorder Examination (EDE), 102
- Eating disorders, 132–133
- Educational and Psychological Measurement*, 10
- Edwards Social Desirability Scale, 697–698
- Emotional support, level of, 695
- Empirical validity, 19
- Equal Opportunity Commission (EEOC), 62
- Equivalence: conceptual, 63; linguistic, 62; metric, 63–66; summary of test, 62; test, and appropriate use of instruments, 61–66
- Essentials of Career Interest Assessment* (Prince and Heiser), 708
- Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association), 50, 54
- Evaluation of Competency to Stand Trial—Revised, 45
- Extraversion (E), 499–500
- Eysenck Personality Questionnaire, 21

F

- Face validity, 19
- Fear of Negative Evaluation Scale (FNE), 132
- Fear Survey for Children, 132
- Fears and anxieties, 132
- Fluid reasoning, 166
- Fluid Reasoning Index (FRI), 146, 147
- Formal Thought Disorder Scale, 105
- Foundations of Behavioral, Social, and Clinical Assessment of Children* (Sattler), 708
- Functional impairment, 684–687

G

- Gambrill Assertion Inventory, 133
- General Behavior Inventory, 425

- General Health Questionnaire (GHQ), 657
- General information, 167, 168
- “General Test Administration Practices Checklist” (Sattler)
- Global Ability Index (GAI), 164
- Global Severity Index (GSI), 657, 660, 691
- Guide to Assessments That Work* (Hunsley and Mash), 68
- Guidelines for Child Custody Evaluations in Family Law Proceedings, 45
- Guidelines for Computer-Based Testing* (Association of Test Publishers), 74
- Guilford-Zimmerman Temperament Survey Restraint scale, 492

H

- Halstead Reitan Neuropsychological Test Battery (HRNTB), 8, 616, 620
- Hamilton Psychiatric Rating Scale for Depression, 663–664
- Handbook of Behavioral Assessments* (Ciminero, Calhoun, & Adams), 116
- Handbook of Diagnostic and Structured Interviewing* (Rogers), 102
- Handbook of Evidence-Based Practice in Clinical Psychology, Child and Adolescent Disorders* (Hersen & Sturmey), 708
- Handbook of Normative Data for Neuropsychological Assessment* (Mitrushina, Boone, Razani, and D’Elia), 617
- Healy Picture Completion, 142
- High S-Anxiety, 669
- High S-Anxiety/Low T-Anxiety, 670
- High T-Anxiety, 669
- High T-Anxiety/Low S-Anxiety, 670
- Hindsight bias, 29

Historical Clinical Risk–, 20, 45
 Holtzman Inkblot Test, 522
 Hopkins Symptom Checklist, 655–656
 Hostility (HOS) symptom dimension, 661
 House-Tree-Person, 7
 Human Figure Drawing, 7
 Hypervigilance Index (HVI), 580

I

Illness Behavior Questionnaire, 67
 Imagery, 134
 Immediate Memory Index, 648, 649
 Impairment, 619
In Our Client's Shoes: Theory and Techniques of Therapeutic Assessment (Finn), 730
 Incremental validity, 24–26
 Individual differences, concept of, 4
 Individual Neuropsychological Testing for Neurotoxicity Battery, 620
International Classification of Disorders (ICD; World Health Organization), 30, 417, 741
 Interpersonal Sensitivity (I-S) symptom dimension, 660–661
 Interpersonal Style Inventory, 492
 Interval sampling, 127
 Interval time sampling, 127
 Intervention options: altering environment, 678–679; and further evaluation, 678; and placement, 678; and self-help, 679; and treatment, 677
 Interview tactics: and avoidance of “why” questions, 92; and comprehensiveness, 92; and concluding interview, 92–93; and directive *versus* nondirective interviews, 90; and nonverbal behaviors, 92; and preliminaries, 89–90; and sequence of interview tactics, 90–92

Inviolacy, 53
 Irrational Belief Test, 132

J

Journal of Behavior Therapy and Experimental Psychiatry, 114, 116
Journal of Behavioral Assessment, 116, 117
Journal of Personality Assessment, 10, 50, 518
Journal of Psychoeducational Assessment, 10

K

K-ABC (Kaufman Assessment Battery for Children), 147
 K-ABC-II (Kaufman Assessment Battery for Children-II), 47, 65, 150, 208
 Kaufman Test of Educational Achievement (KTEA), 208
 Koh's Block Design, 142
 Koppitz Developmental Scoring System, 636

L

Language Index, 650
 Lawrence Psychological-Forensic Examination, 92
 Lazarus BASIC-ID, 80, 120
 Leiter International Performance Scale, 64
 Lexical knowledge, 167
 Linguistic equivalence, 62
 Long-term memory, 167, 168
 Luria-Nebraska Neuropsychological Test Battery, 8, 617, 619, 620
 Luria's Pass model (Planning-Attention-Successive-Sequencing), 153

M

MacArthur Competence Assessment Tool, 45
 Magnification/minimization, 131–132

- Manifest Anxiety Scale, 667, 668
 Marital Attitude Survey, 133
 Marital relationships, 133
 Mayo Clinic, 272
 MBTI. *See* Myers Briggs Type Indicator
 MCMI. *See* Millon Clinical Multiaxial Inventory
Measures for Clinical Practice: Sourcebook (Fischer and Corcoran), 10, 68
 Members of the Personality Profiles of Cultures Project, 489
 Memory Assessment Scales, 629
Mental Measurement Database, 68
Mental Measurements Yearbook (Carlson, Geisinger, & Jonson), 10, 68, 272, 490
 Mental status evaluation, 93–99; and feeling (affect and mood), 97; and format for mental status and history, 95–96 Fig. 3.1; and general appearance, behavior, and relatedness, 94–97; and perception and thinking, 97–99; speech and language in, 97
 Michigan Alcoholism Screening Test, 425
 Millon Adolescent Clinical Inventory (MACI), 47
 Millon Behavioral Health Inventory, 67
 Millon Clinical Multiaxial Inventories (HCMI), 7, 56, 67, 492; assets and limitations of, 427–431; and clinical personality patterns, 439–476; and clinical syndromes, 486–487; Debasement Index (Z), 438–439; Desirability Index (Y), 438; and determining profile validity, 433; and development of MCMI-II, 421; and development of MCMI-III, 421; and development of MCMI-IV, 421–422; Disclosure Index (X), 438; and elaborating on treatment implications and recommendations, 436; history and development of, 419–424; Inconsistency Index (Scale W), 437–438; interpretation procedure for, 432–436; and interpreting clinical syndrome scales, 435; and interpreting Personality Disorder scales, 434; Invalidity Index (Scale V), 437; and providing diagnostic impressions, 436; recommended reading for, 488; reliability and validity of, 424–426; and reviewing noteworthy responses, 435–436; scale categories, abbreviation, number of items, and reliabilities, 418 Tab. 9.1; and severe personality pathology, 476–486; and severe syndromes, 488; theoretical considerations of, 422–424; use of, with diverse groups, 431–432; validity scales in, 436–439
Millon Clinical Multiaxial Inventory–IV Manual (Millon, Grossman, and Millon), 418
 Millon-Illinois Self-Report Inventory (MI-SRI), 419
 Mini Mental State Examination, 94, 624, 646
 Minnesota Multiphasic Personality Inventory (MMPI), 7, 114, 371, 492, 524; administration of, 262–263; and age, 259; assets and limitations of, 255–258; and content nonresponsiveness scales (CNS), 348–349; and ethnicity, 259–262; and F (Infrequency) scale, 275; and fake bad scale (FBS), 277; and Fb (F back) scale, 276; and Fp (Infrequency-Psychopathology) scale, 276–277; history and development of, 246–251; and interest scales, 366; and K (Correction) scale, 278–279; and L (Lie) scale, 277–278; and over-reporting

- scales, 349–352; overview of, 243–246; point codes, 302–328; recommended reading for, 370; reliability and validity of, 251–255; and S (Superlative) scale, 279; six goals of, 249; and thought dysfunction (THD), 354–355; and TRIN (true response inconsistency scale), 274–275; and underreporting scales, 352–353; use of, with diverse groups, 259–262; validity scales, 273–274; and VRIN (variable inconsistency scale), 274
- Minnesota Multiphasic Personality Inventory-2 (MMPI-2), 56, 66, 249, 250; clinical scales, 279–302; computerized interpretation, 272–273; content scales, 328–331; critical items, 337–338; Harris-Lingoes and SI subscales, 334–337; higher-order scales, 353–354; interpretation procedure, 263–272; and MMPI–A supplementary scales, 337–341; RF interpretation procedure, 341–346; RF personality psychopathology five scales, 366–369; RF restructured clinical scales, 355–361; RF specific problem scales, 361–366; RF validity scales, 346–348
- Minnesota Multiphasic Personality Inventory-A (MMPI-A), 47; content scales, 331–334; supplementary scales, 341
- Missouri Automated Mental Status Examination Checklist, 94
- MMPI. *See* Minnesota Multiphasic Personality Inventory
- MMPI/MMPI— 2, 65, 250
- Modern Psychopathology* (Millon), 419
- Motivation Assessment Scale (MAS), 129
- Multilingual Aphasia Examination, 626
- Myers Briggs Type Indicator (MBTI), 21
- N
- National Adult Reading Test (NART), 201
- National Council for Measurement in Education, 18
- NEO Personality Inventory: assets and limitations of, 493–494; history and development of, 490–491; interpretation procedure, 495–512; and personality styles, 505–512; recommended reading for, 512; reliability and validity of, 491–493; use of, with diverse groups, 494–495
- NEO-PI-3. *See* Neo Personality Inventory
- NEO-PI-R. *See* Neo Personality Inventory
- Neurobehavioral Assessment Format, 623–624
- Neurobehavioral Evaluation System (NES), 72
- Neuropsychological Assessment* (Lezak and others), 68, 708
- Neuropsychological Assessment Battery, 616
- Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration* (Groth-Marnat), 615, 708
- Neuropsychological functioning, domains of, 624–633; and attention, 624–625; and executive functions, 630–633; and frequently used tests to measure, 624 Tab. 12.2; and language, 625–627; and memory, 627–629; and visuospatial functions, 629–630
- Neuropsychological History Questionnaire, 623
- Neuropsychological impairment: and Bender Visual-Motor Gestalt Test, second edition, 633; and domains of neuropsychological functioning, 624; history and

Neuropsychological impairment

(*Continued*)

development of screening for,

617–618; interviewing for,

617–618; screening for, 615–617

Neuropsychological Status Examination,

92, 623

Neuropsychological Symptom Checklist,

620, 623

Neuroticism (N), 497–499

NIMH Core Neuropsychological

Battery, 620

Nonverbal fluid reasoning, 167

North Carolina Mental Status

Examination (Ruegg, Ekstrom,

Evans, & Golden), 94

Northern Plains Bicultural Immersion

Scale, 61

O

Obsessive Style Index (OBS), 580–581

Obsessive-Compulsive (O-C) symptom
dimension, 660

Office of Contract Compliance, 62

Office of Strategic Services Staff, 115

Oklahoma Premorbid Intelligence

Estimate (OPIE-4), 201

Openness (O) scale, 500–502

Overgeneralization, 131–132

P

Padua Inventory, 668

PAI RXR (Treatment Rejection) scale,
698

Paranoid Ideation (PAR) symptom

dimension, 661

Peabody Picture Vocabulary Test–IV, 64,

626–627

Pearson Assessments, 272

Perceptual Reasoning Index (PRI), 146

Perceptual Thinking Index (PTI), 521,
579

Personality Assessment Inventory (PAI),
7, 492; additional clusters,

410–415; assets and limitations of,
375–377; clinical scales, 384–404;
critical items, 415–416; and
environmental perception,
412–414; history and development
of, 373; interpersonal scales,
408–410; and interpersonal style,
411–412; interpretation
procedure, 377–380; overview of,
371–372; and potential for
dangerousness, 414–415;
recommended reading, 416;
reliability and validity of,
373–375; and self-concept, 410;
treatment scales, 404–408; use of,
with diverse groups, 377; and
validity, clinical, treatment, and
interpersonal personality
assessment inventory scales, 372
Tab. 8.1; validity scales in,
380–384

Personality Assessment Inventory,
clinical scales: and alcohol
problems, 401–402; and antisocial
features (ANT), 400–401; and
anxiety (ANX), 386–388; and
anxiety-related disorders (ARD),
388–390; and borderline features
(BOR), 397–399; and depression
(DEP), 390–392; and drug
problems (DRG), 403–404; and
mania (MAN), 392–394; and
Paranoia (PAR), 394–395; and
schizophrenia (SCZ), 395–396;
and somatic complaints (SOM),
384–386

Personality Assessment Inventory,
interpersonal scales: and
dominance (DOM), 408–409; and
warmth (WRM), 409–410

Personality Assessment Inventory,
interpretation procedure: and
analysis of critical terms, 379; and
configural interpretation, 380; and
full scale interpretation, 379; and
interpreting test validity, 378–379;

- and subscale interpretation, 379–380
- Personality Assessment Inventory, treatment scales: and aggression (AGG), 404–405; and nonsupport (NON), 407; and stress (STR), 406–407; and suicidal ideation (SUI), 405–406; and treatment rejection (RXR), 407–408
- Personality Assessment Inventory, validity scales: and inconsistency (INC), 380–381; and infrequency (INF), 381; and negative impression (NIM, MAL, RDF), 381–382; and positive impression (PIM, DEF, CDF, ALC Est, DRG Est), 382–384
- Personality Assessment Inventory Professional Manual* (Morey, 372, 374)
- Personality Assessment Proficiency Report Review Form, 709–715 Fig. 15.1
- Personality Guided Therapy* (Millon), 436
- Personality Inventory for Children–2 (PIC–2), 47
- Pinther-Paterson Test, 142
- Pittsburgh Occupational Exposure Test, 620
- Pocket to Clinical Psychiatry* (Sadock & Sadock), 707–708
- Positive Symptom Total (PST), 660
- Predictive validity, 19, 20
- “Privacy and Behavioral Research” (Office of Science and Technology, 50
- Problem complexity and chronicity, 687–690
- Problem context: and coping style, 692–694; and current life circumstances, 696; and social support, 694–696; understanding, 692–694
- Problem Oriented Record Week, 116
- Problem solving, without visual-motor speed, 167
- Psych Screen, 272
- Psychiatric Diagnosis* (Woodruff, Goodwin, & Guze), 101
- Psychiatric Diagnostic Interview, 109
- Psychoactive Substance Use Disorders module, 104
- Psychodiagnostik* (Rorschach), 514
- Psychological Assessment Resources, 272
- Psychological Corporation, 217, 219, 627
- Psychological Report: addressing diversity in, 722; and background information/relevant history, 732–735; and behavioral observations, 735–737; and case-focused reports, 723; and comprehensiveness, 708, 716; deciding what to include in, 716; diagnosis and recommendations in, 722–723; emphasis in, 720–722; and evaluation procedures, 732; and examples of general topics around which case presentation may be conceptualized, 717 Tab. 15.1; and feedback, 728–729; format for, 730–731; general guidelines for, 708; and individualization, 723–724; interpretation and impressions in, 737–741; overview, 707–708; and Personality Assessment Proficiency Report form, 709–715 Fig. 15.1; and presenting test interpretations, 719–720; and referral question, 731–732; samples of, 743–772; and style, 724–726; summary, 718; summary and recommendations, 741–742; terminology in, 726–727; topics, 716; use of raw data in, 727–728
- Psychological Report* (Klopfer), 738
- Psychological Report Writing* (Tallent), 704, 717 Tab. 15.1

Psychological Report Writing Assistant (Groth-Marnat and Davis), 724

Psychological tests, evaluating: practical considerations for, 10–11; and reliability, 12–14; standardization in, 11–12; theoretical orientation for, 10

Psychometric Software, 272

Psychoticism (PSY) symptom dimension, 661–662

Publication Manual of the American Psychological Association (Hollis and Donna), 725–726

Q

Quantitative reasoning, 167

R

Rathus Assertiveness Schedule (RAS), 119, 120, 133; modified (MRAS), 119

Rational Belief Inventory, 132

Raven's Progressive Matrices, 64–65

RBANS. *See* Repeatable Battery for the Assessment of Neuropsychological Status Update (RBANS)

Relationship Attribution Measure, 133

Relationships Beliefs Inventory, 133

Reliability, 12–14; and alternate forms, 14–15; and construct validity, 21–23; and content validity, 18–19; and criterion validity, 19–21; and internal consistency, 15–16; interscorer, 16; selecting forms of, 16–17; test-retest, 14; and validity, 17–18

Renard Diagnostic Interview, 8, 101, 110

Repeatable Battery for the Assessment of Neuropsychological Status Update (RBANS), 148, 149, 615, 616; administration of, 650; assets and limitations of, 648–650; description of, 647 Tab. 12.4; history and development of, 645–648; interpretation guidelines

for, 650–652; recommended reading for, 652; reliability and validity in, 648–649

Research Diagnostic Criteria (RDC), 101

Revised California Personality Inventory Self-Control scale, 492

Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery (Heaton, Miller, Taylor, and Grant), 617

Rey Auditory-Verbal Learning Test, 618, 629

Rey-Osterrieth Complex Figure Test, 618, 629, 648

Rivermead Behavioral Memory Test, 629

Rogers Criminal Responsibility Assessment Scales, 45

Rorschach system: assets and limitations of, 523–527; background, 513–514; history and development of, 514–518; use of, with diverse groups, 527

Rorschach: A Comprehensive System, 516

Rorschach Performance Assessment System (R-PAS), 7, 689; administration, 581–584; administration behaviors and observations, 593–594; and clarification phase, 583–584; coding, 584–592, 584–593; cognitive codes, 589–591; content codes, 585, 586; determinants codes, 588–589; engagement and cognitive processing: page 1, 594–598; engagement and cognitive processing: page 2, 605–608; and establishing rapport, 581; form quality, 586–587; and giving test instructions, 581–582; interpretation, 593–602; location and space codes, 584–585; object qualities codes, 585, 587; and Page 1 Self and Other Representation, 602–605; and Page 2 Engagement and Cognitive Processing, 605–611; and Page 2 Self and

- Other Representations, 611–614; perception and thinking problems: page 2, 608–609; perception and thinking problems: page 1, 598–601; popular responses, 587–588; recommended reading for, 614; and response phase, 582–583; scoring structural summary of, 592–593; self and other representation: page 1, 602–605; self and other representation: page 2, 611–614; stress and distress: page 1, 601–602; stress and distress: page 2, 609–611; thematic codes, 591–592
- Rorschach Performance Assessment System Administration, Coding, Interpretation, and Technical Manual* (Meyer and others), 584
- Rorschach Research Council (RRC), 517–518
- S**
- Satz-Mogel approach, 208, 209
- SCALD profile, 205
- Schedule for Affective Disorders and Schizophrenia (SADS), 77, 85, 101, 104–107; adult version, 105–106; child version (K-SADS), 106–107
- Scholastic Aptitude Test (SAY), 25
- Selective abstraction, 131–132
- Self-efficacy, 133
- Self-Help That Works* (Norcross *et al.*), 742
- SEM. *See* Standard error of measurement
- Sensorium, 98
- Severe Personality Pathology (Millon): Scale C: Unstable-Borderline-Cyclophrenic (UBCycloph), 479–482; Scale P: Mistrustful-Paranoid-Paraphrenic (MPParaph), 483–486; Scale S: Eccentric-Schizotypal-Schizophrenic (ESSchizoph), 477–479
- Severe Syndromes: Scale CC: Major Depression, 48; Scale PP: Delusional, 48; Scale SS: Schizophrenic Spectrum, 48
- Shedler-Westen Assessment Procedure–200 (SWAP–200), 428
- Short-term memory, 167
- Social Avoidance and Distress Scale, 132
- Social Interaction Self-Statement Test, 132
- Social skills, 133
- Social Support Questionnaire, 695
- Society for Personality Assessment (SPA), 708, 709
- Somatization (SOM), 660
- Specific Relationship Standards, 133
- Standard error of measurement (SEM), 17
- Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education), 18, 50
- Standards for Educational and Psychological Tests* (American Psychological Association), 50
- Stanford Binet, 65, 208
- Stanford Binet–V, 47, 142, 208
- Stanford Shyness Survey, 133
- State of Change Readiness and Treatment Eagerness Scale (SOCRATES), 700
- State-Trait Anxiety Inventory (STAI), 685, 691; interpretation in, 669–670; overview, 666–667; reliability and validity of, 667–669; use of, with diverse groups, 669
- State-Trait Anxiety Inventory for Children (STAIC), 667

- Static 99, 45
- Structured Clinical Interview for the DSM (SCID), 8, 77, 102–104
- Structured Interview for *DSM-IV* Personality (SCID-IV), 741
- Structured interviews, 100–111; and Diagnostic Interview Schedule (DIS), 107; frequently used, by categories, 103 Tab. 3.2
- STS Clinician's Rating Form* (Corbella), 704
- Subjective distress, 690–692
- Suicide Constellation (S-CON), 580
- Survey of Homosexual Interactions, 133
- Symptom Checklist-90-R and Brief Symptom Inventory, 425, 691; interpreting, 659–662; reliability and validity of, 657–659; use of, with diverse populations, 659
- Symptom Checklist-90-R and Brief Symptom Inventory (BSI): reliability and validity of, 657–659
- Symptom level/item, 662
- Synthetic validity, 20
- Systematic Treatment Selection (STS)/Innerlife Approach, 655, 703–705; characteristics included in treatment planning, 704 Tab. 14.1; and STS characteristics included in treatment planning, 704 Tab. 14.1
- T**
- Taylor Manifest Anxiety Scale, 697
- Technical and Interpretive Manual* (Wechsler), 150
- Test of Nonverbal Abilities (TONI-4), 64–65, 155
- Test publishers/distributors, 773–775 (Appendix A)
- Testing bias, 61
- Testing organizations, 777–778 (Appendix B)
- Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business* (Maddox), 10, 68
- Tests in Print* (Murphy, Geisinger, Carlson, & Spies), 10
- Tests in Print VIII* (Murphy, Geisinger, Carlson, Spies), 68
- Thematic Apperception Test (TAT), 7, 631, 689, 721, 738
- Therapeutic Feedback with the MMPI-2: A Positive Psychology Approach* (Levak, Siegel, and Nichols), 730
- Time sampling, 127
- Trail Making Test (Army Individual Test), 632; Part A, 632; Part B, 633
- Trait psychology, concept of, 4
- Treatment planning, 671–672; and case formulation, 680–683; and clinical decision making, 679–680; development and approaches to, 672–677; intervention options in, 677–679
- Treatment Readiness Scale, 697
- Treatment recommendations, types of, 742 Tab. 15.3
- Treatment-specific client characteristics: and resistance, 696–699; and stage of change, 699–701
- U**
- Unified Tri-service Cognitive Performance Assessment Battery, 73
- Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB), 73
- United States Census Bureau, 59
- Universal Nonverbal Intelligence Test-2 (UNIT-2), 64–65, 155
- University of Minnesota, 246
- University of Rhode Island Change Assessment (URICA), 700

V

- Validity, in clinical practice, 23; conceptual, 26; incremental, 24–26
- Verbal Fluid Reasoning, 167
- Vineland Adaptive Behavior Scales–II, 47, 208
- Visual processing, 166–167
- Visual Spatial Index (VSI), 146
- Visual-motor speed, 167
- Visuospatial/Constructional Index, 648, 650

W

- WAIS-IV Administration and Scoring Manual*, 156, 158, 161, 163, 164, 166, 208
- WAIS-IV Technical and Interpretive Manual* (Wechsler), 150, 164, 181, 194, 218
- WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson), 144, 173
- Wechsler Adult Intelligence Scale (WAIS), 142, 626–627
- Wechsler Adult Intelligence Scale, WAIS–IV/WISC–V, 66, 139, 140, 629, 630; intelligence classifications, 159 Tab. 5.5; level I: full scale IQ (and Global Ability Index), 164–165; level II: indexes and additional groupings, 165–168; level III: interpreting subtest variability, 168–169; level IV: qualitative/process analysis, 170–173; level V: intrasubtest variability, 173; successive-level interpretation procedure for, 161–164; and summary of successive five-level interpretive procedures of, 162 Tab. 5.6
- Wechsler Adult Intelligence Scale, revised (WAIS-R), 142
- Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV), 47, 143, 144, 146, 219; major changes on, 143

Table 5.1; organization of, subtests, 144 Table 5.2, 147; reliability and validity of, 147–150

- Wechsler Adult Intelligence Scale-Third Edition (WAIS-III), 142, 143, 648
- Wechsler indexes and subtests, 173–196; and picture completion, 181–182; and arithmetic, 185, 189–190; and block design subtest, 179–181; and cancellation, 193–194; and Coding, 191–192; and complementary Indexes, 194; and comprehension subtest, 177–178; and Digit Span, 186–188; and figure weights subtest, 183–184; and fluid reasoning, 182–183; and information subtest, 176–177; and Letter-Number Sequencing, 190; and matrix reasoning, 183; and naming speed, 195; and perceptual reasoning, 178–179, 182–183; and picture concepts, 184–185; and picture span, 188–189; and Processing Speed Index, 190–191; and similarities subtest, 175; and Symbol Search, 193; and Symbol Translation, 195–196; and Verbal Comprehension Index/subtests, 174–196; and visual puzzles subtest, 181; and Working Memory Index, 185–186
- Wechsler Individual Achievement Test-Second Edition (WIAT-II), 143, 144, 636
- Wechsler Individual Achievement Test-Third Edition (WIAT-III), 47, 143, 200, 298, 627
- Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V), 143, 146, 208; major changes on, 143 Tab. 5.1; reliability and validity of, 150–151
- Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV), 143–144;

- Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) (*Continued*)
organization of indexes and subtests for, 146
- Wechsler Intelligence Scale for Children-Revised (WISC-R), 145
- Wechsler Intelligence Scale for Children-Third Edition (WISC-III), 145, 636
- Wechsler Intelligence Scale for Children-V, 47
- Wechsler intelligence scales: 202, 210;
and Alzheimer's disease, 203–204;
and assessing brain damage, 196–199; assets and limitations of, 151–153; and best four-subtest short form, 211; and best two- and three-subtest short forms, 211;
cautions and guidelines in administration of, 159–161; and estimating premorbid IQ, 200–203; and gifted children, 208–209; history and development of, 142–147; and intellectual disability, 207–208; and learning disabilities, 204–207; and meaning of IQ scores, 156–159; and pros and cons of testing intelligence, 139–142; recommended reading for 213, 212–213; relationship of Wechsler scores to various types of standard measures, 158 Fig. 5.1; and Satz-Mogel/Yudin and modified formats, 212–213; and seven-subtest short forms, 211–212; short forms, 209–210; use of, with diverse groups, 153–156; and Wechsler Abbreviated Scale of Intelligence (WASI), 210; and Wechsler indexes and subtests, 173–196
- Wechsler Memory Scale (WMS), 7, 215, 629, 649; assets and limitations of, 224–226; and Auditory Memory Index (AMI), 231–232; and Auditory Memory Index *versus* Visual Memory Index, 233–234; and comparing scores of WMS-IV with WAIS-IV, 238–240; and Delayed Memory Index, 236–237; and General Ability Index *versus* Auditory Memory Index, 240; and General Ability Index *versus* Visual Memory Index, 240; and General Ability Index *versus* Visual Working Memory Index, 240; history and development of, 216; and Immediate Memory Index, 236; and Immediate Memory Index *versus* Delayed Memory Index contrast scaled score, 237–238; interpretation procedure for, 227–230; interpreting patterns of index scores in, 230–231; and malingering and evaluating change, 241–242; recommended reading for, 242; reliability and validity of, 222–224; use of, with diverse groups, 226–227; and Visual Memory Index, 232–233, 235; and Visual Memory Index *versus* Visual Memory Index Contrast Scaled Score, 235–236; and Visual Working Memory Index (VWMI), 234–235
- Wechsler Memory Scale-Fourth Edition (WMS-IV), 143, 144, 624, 629; adult battery indexes, primary subtests, 221 Tab. 6.2; older adult battery, 222 Tab. 6.3
- Wechsler Memory Scale-Revised (WMS-R), 217, 218
- Wechsler Memory Scale-Third Edition (WMS-III), 218, 219
- Wechsler Nonverbal Scale of Ability (WNV), 155
- Wechsler Preschool and Primary Scale of Intelligence (WPPSI), 147
- Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R), 147

- Wechsler Test of Adult Reading (WTAR), 201
- Wechsler Working Memory Index, 631
- Wechsler-Bellevue Intelligence Scale, 142, 144
- Welsh Anxiety Scale, 667
- Western Psychological Services, 272
- Wide Range Achievement Test (WRAT-IV), 8, 47, 200, 208, 627
- Wide Range Assessment of Memory and Learning–II (WRAML–2), 629
- Wolpe-Lazarus Assertion Inventory, 133
- Woodcock-Johnson Psychoeducational Battery–IV, 47
- Woodcock-Johnson Tests of Achievement–III, 147, 627
- Woodworth Personal Data Sheet, 655–656
- Work Progress Administration, 247
- World Health Organization, 30
- Writing to Clients and Referring Professionals about Psychological Assessment Results: A Handbook of Style and Grammar* (Allyn), 724