ROUTLEDGE

HANDBOOK ON ASSESSMENTS FOR GIFTED LEARNERS

IDENTIFICATION, LEARNING PROGRESS, AND EVALUATION

EDITED BY
Susan K. Johnsen and Joyce VanTassel-Baska

A **Prufrock Press** Book

Handbook on Assessments for Gifted Learners

This essential handbook is a comprehensive and systematic examination of the assessment of gifted and advanced students and their programs, and a must-have resource for coordinators and directors at state and local levels.

Handbook on Assessments for Gifted Learners explores issues associated with building an effective identification system, clarifies and interprets the need for targeted learning progress assessments for gifted learners, and discusses program evaluation, assessments, and processes used to gauge programs' success. Engaging chapters written by both academic and practitioner experts provide research-based, practical ideas for identifying and measuring the progress of gifted and advanced learners. Readers will benefit from informed recommendations stemming from current research conducted specifically for this text.

Susan K. Johnsen is Professor Emerita of Educational Psychology at Baylor University and Editor-in-Chief of *Gifted Child Today*.

Joyce VanTassel-Baska is the Smith Professor Emerita of Education and former Executive Director of the Center for Gifted Education at The College of William and Mary in Virginia.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Handbook on Assessments for Gifted Learners

Identification, Learning Progress, and Evaluation

Edited by Susan K. Johnsen and Joyce VanTassel-Baska

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Cover image: © Getty Images

First published 2022
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

ISBN: 978-1-032-25987-1 (hbk)
ISBN: 978-1-032-25984-0 (pbk)
ISBN: 978-1-003-28599-1 (ebk)

DOI: 10.4324/9781003285991

Typeset in Bembo
by Newgen Publishing UK

Contents

Author Biographies

Co-Editors

Susan K. Johnsen, Ph.D., is Professor Emerita of Educational Psychology at Baylor University. Dr. Johnsen is editor-in-chief of *Gifted Child Today* and has published 30 books and more than 350 articles, monographs, technical reports, and chapters related to gifted education. Included among her books are *Identifying Gifted Students: A Practical Guide* (2018), *Implementing Research-Based Gifted Education Practices* (2021), *Classroom Management for Gifted and Twice- Exceptional Students Using Functional Behavior Assessment* (2021), and *NAGC Pre-K-Grade 12 Gifted Education Programming Standards: A Guide to Planning and Implementing High-Quality Services* (2022). She has written three tests used in identifying gifted students: Test of Mathematical Abilities for Gifted Students (TOMAGS-2), Test of Nonverbal Intelligence (TONI-4) and Screening Assessment Gifted Students (SAGES-3). She is past president of The Association for the Gifted (TAG), Council for Exceptional Children, and past president of the Texas Association for Gifted and Talented (TAGT). She has received awards for her work in the field of education, including NAGC's Ann Isaac's Award, NAGC's President's Award, and CEC's Leadership Award.

Joyce VanTassel-Baska, Ed.D., is the Jody and Layton Smith Professor Emerita of Education and founding director of the Center for Gifted Education at The College of William and Mary in Virginia where she developed a graduate program and a research and development center in gifted education. She also founded the Midwest Talent Search and the Center for Talent Development at Northwestern University. Dr. VanTassel-Baska has published widely including 33 books and over 700 refereed journal articles, book chapters, and scholarly reports. Her most recent books are: *Talent Development: Theory, Research, and Practice* (Ed.) (2022) and *Curriculum Planning and Instructional Design for Gifted Learners* (3rd edition) (2020) (with Ariel Baska). She has received numerous awards for her work, including scholar and service awards from the National Association for Gifted Children, Council for Exceptional Children – The Association of the Gifted, Mensa, the American Educational Research Association, and the World Council on Gifted and Talented Education. She was selected as a Fulbright Scholar to New Zealand in 2000 and a Visiting Scholar to Cambridge University in England in 1993. She served as President of both CEC-TAG and NAGC.

Authors

Cheryll M. Adams, Ph.D., received her doctorate from the University of Virginia and is the Director Emerita of the Center for Gifted Studies and Talent Development, Ball State University. She was a classroom teacher for 15 years. She currently teaches on-line courses in gifted education and has presented widely in the field. She has authored or coauthored numerous publications in professional journals, as well as several books and book chapters. She co-authored, received, directed, and was a principal investigator on three Javits grants focused on underrepresented populations in gifted education. She is a former member of the Board of Directors of NAGC, a Past President of The Association for the Gifted (CEC) and the Indiana Association for the Gifted. She is a past Chair of the NAGC Professional Standards Committee and the Research and Evaluation Network. She has received the Early Leader Award, Distinguished Service Award, and President's Award from NAGC.

Cecelia A. Boswell, Ed.D., De Leon, TX, was an Advanced Academics consultant for an Education Service Center and an urban school district. She is an independent consultant, Austin Creek Education Systems. She has provided multiple products for the Texas Education Agency, Texas Association for the Gifted and Talented (TAGT), co-authored five books on gifted education and has co-edited two books on twice exceptional learners and culturally responsive teaching. She has been board member and President of TAGT and CEC-TAG. Her research interests include gifted, 2e, and rural gifted education.

Catherine M. Brighton, Ph.D., is Senior Associate Dean for Academic Programs & Student Services and Professor of Curriculum & Instruction in the School of Education and Human Development, University of Virginia, where she also serves as the co-director of the University of Virginia Institutes on Academic Diversity. Catherine's research interests include teacher change and school reform initiatives, differentiating curriculum, instruction, and assessment, and qualitative methodologies. She has served as a Co-Principal Investigator on a US Department of Education-funded grant on talent development through the lens of literacy. Catherine also led a multi-year project for the National Research Center on the Gifted & Talented focus on differentiating instruction in middle-grade classrooms with diverse student populations.

Elissa F. Brown, Ph.D., is Director of the Binghamton University Community Schools, Regional Network. From 2012 to 2021, she was the director of the Hunter College Center for Gifted Studies in New York City and served as a distinguished lecturer and program leader of Hunter's Advanced Certificate Program in Gifted Education. Elissa, formerly, was the North Carolina state director of gifted education and the director of the Center for Gifted Education at the College of William & Mary. She has been a district gifted program coordinator, the principal of a specialized high school, and a teacher of gifted learners at elementary, middle, and high school levels. She has served as an adjunct professor at several universities, including Rutgers, Vanderbilt, and Duke University. She is a published author in the field of gifted education and presents widely. She currently serves as interim co-president of GiftedNYS, a New York statewide advocacy association for gifted education.

Carolyn M. Callahan, Ph.D., is the Commonwealth Professor of Education Emeritus at the University of Virginia. Dr. Callahan has been principal investigator on projects of the National Center for Research on Gifted Education (formerly NRC/GT) and multiple Javits grants. Her work in these grants has focused on curriculum development, implementation, and evaluation. She has been recognized as Outstanding Professor of the Commonwealth of Virginia and Distinguished Scholar of the National Association for Gifted Children and has served as Editor of *Gifted Child Quarterly*. Dr. Callahan has published over 275 articles and 60 book chapters.

Melanie Caughey, Ph.D., received her doctoral degree in gifted education from the University of Virginia, where she served as the academic coordinator for the Saturday and Summer Enrichment Program (SEP) and as a graduate research assistant on the Javits grant *Promoting PLACE in Rural Schools*. Since graduating, she has taught undergraduate coursework on ELA methods and secondary content literacy, as well as graduate coursework on instructional coaching, gifted education, curriculum development, differentiated instruction, creativity, social-emotional needs of gifted learners, development of programming and services, action research, and introductory research methods. She has presented her work at national, state, and local conferences and published several journals articles and book chapters. She has served on the Gifted Advisory Council for the Ohio Department of Education and has been active in multiple gifted education organizations.

Marla Caviness-French, M.Ed. from Central Washington University, MLIS from University of Colorado at Denver. Before becoming a Gifted and Talented Resource Teacher in the Jeffco Public Schools, Golden, CO., Marla taught social studies and language arts at the secondary level for 14 years and was a teacher librarian for nine years. Marla is also the CogAT coordinator for Jeffco. Marla has presented various GT topics at the Colorado Association for Gifted and Talented (CAGT), Colorado Association for Libraries (CalCON) Association of Career and Technical Education (ACTE), and NAGC in order to advocate for the use of cognitive data to guide instruction and needs of secondary gifted students.

Mary M. Christopher, Ph.D., comes with a wealth of information to support parents, teachers, and administrators working with gifted children. She earned a Ph.D. in curriculum and instruction with a concentration in gifted education from Texas Tech University. After teaching gifted children in public school, she worked for 25 years as a Professor in Gifted Education and Leadership at Hardin-Simmons University. Currently, she works as consultant/owner of CGL Consulting, LLC, providing professional development, leadership development, parent support, and program evaluation for individuals, districts, and other entities throughout Texas and the United States. She is Past-President of the Texas Association for Gifted and Talented. Dr. Christopher has written many articles and presented throughout Texas and the United States. Her most recent book is *Leadership for Kids: Curriculum for Building Intentional Leadership in Gifted Children*.

Yara N. Farah, Ph.D., has been working for the past 13 years in the field of education. She has a doctorate in educational psychology from Baylor University, and an MA in special education from California State University, Northridge. Her expertise in the field also stems from the numerous diverse pedagogical positions she has assumed: working as a homeroom teacher in International Baccalaureate programs, being a special education teacher in elementary school, teaching in education psychology departments at various higher education institutions, and conducting applied research in education. Dr. Farah has more than 12 publications in peer-reviewed journals and has conducted numerous professional presentations at national and international conventions. Since 2012, Dr. Farah has also participated in educational consulting, and helped lead several academia-related workshops and trainings.

Maryann R. Hebda, M.S., is an adjunct instructor and doctoral student in Educational Psychology at Baylor University, specializing in Gifted and Talented Education. Her most recent research, conference presentations, and publications feature talent development, achievement motivation in gifted learners, and twice-exceptionality. She is an active member of multiple committees in the American Educational Research Association and has received two first place graduate student research awards from the National Association for Gifted Children's Research and Evaluation Network. Prior to relocating to Texas, Hebda taught STEM, special education, and gifted education in Nebraska and Kansas for 13 years. She received her M.S. in Special Education-Gifted Emphasis from Emporia State University and B.S. in Elementary and Special Education from Nebraska Wesleyan University.

Michelle Hock, M.Ed., is a doctoral candidate in curriculum and instruction with a focus area in gifted education at the University of Virginia's School of Education and Human Development. She holds a Master of Education degree in curriculum and instruction with a gifted education emphasis. Michelle currently serves as a graduate instructor and teaching assistant in the secondary Master of Teaching program at the University of Virginia. Her interests include pre-service teacher education, secondary English education, gifted education, and classroom assessment. Prior to pursuing her Ed.D., Michelle taught secondary English/language arts courses and journalism courses in public schools in Pennsylvania. She also taught English/language arts courses designed specifically for gifted and talented learners.

Kelly L. Kearney, Ph.D., completed her degree in Educational Psychology with emphasis in Gifted and Talented Education at the University of Connecticut in 2014. She worked on Project SPARK and Project LIFT, two Javits-funded initiatives focused on working with schools and teachers to recognize and respond to advanced academic potential in the early grades, particularly with students from underserved populations. Previously Dr. Kearney was a first grade teacher and an assistant to enrichment coordinator and worked with teachers and students in school-based and out-of-school programs. Her research has focused on resilience in advanced learners from underserved populations.

Todd Kettler, Ph.D., is Associate Professor of Educational Psychology in the School of Education at Baylor University where he is also an affiliated faculty member of Baylor's Center for Gifted Education and Talent Development. He coordinates graduate programs specializing in gifted and talented education for the Department of Educational Psychology, and he directs the Baylor Exceptionalities Research Group (BERG). Dr. Kettler is a previous winner of a Mensa Research Award, and he won the Legacy Award for the best scholarly book in gifted education for his work, *Modern Curriculum for Gifted and Advanced Academic Students*. He has served as the co-editor of the *Journal for Advanced Academics* and the *Journal of Gifted Education and Creativity*. His current work focuses on the pedagogies of talent development including his most recent book *Personalized Learning in Gifted Education*.

Joni Lakin, Ph.D., The University of Iowa, is an Associate Professor of Educational Studies at the University of Alabama and affiliated with the Gifted Education and Talent Development Office (GETDO). She is lead author of the Cognitive Abilities Test (Form 8) and studies test validity and fairness with a particular interest in the equitable identification of students with academic talents. Her work in this area received the 2019 NAGC "*Gifted Child Quarterly* Paper of the Year Award." Her work on assessing individual differences informs educational practices by promoting talent development among marginalized communities. Her federally funded research on developing new spatial assessments seeks to engage the talents of students with potential to excel in science, technology, engineering, and mathematics (STEM) fields.

Kristen N. Lamb, Ph.D., is an Assistant Professor in the College of Education at the University of Alabama. She earned her doctorate in Educational Psychology with an emphasis in gifted education and creativity from the University of North Texas and previously served her community as a K-12 public educator. She studies equity issues in advanced academics, the role of creativity in talent development, and classroom conditions conducive to developing creative thinking and advanced academic achievement. Dr. Lamb is the co-author of *Developing Creativity in the Classroom: Learning and Innovation for 21st-Century Schools* and has published a number of articles and invited book chapters on the topics of creativity and gifted education. Her work in gifted education and creativity has received scholarly recognition, including the 2019 Paper of the Year for *Gifted Child Quarterly*, 5th annual Paper of the Year for *Journal for the Education of the Gifted,* and the 1st Place Dissertation Award from the National Association for Gifted Children. She is currently an associate editor for the *Journal for Education of the Gifted* and serves on the executive board for the Alabama Association for Gifted Children.

Lindsay Ellis Lee, Ph.D., is a Research Assistant Professor in the Department of Pediatrics at East Tennessee State University. She graduated from the University of North Texas (UNT) with a Ph.D. in Educational Psychology with a concentration in Gifted and Talented Education. Lindsay's research interests include evaluating systemic factors that relate to the development of talent, inclusive of the instrumentation and processes involved in identifying students for advanced learning opportunities. Her dissertation focused on evaluating program diversity and the probability of gifted identification with the Torrance Test of Creative Thinking. She is currently working on optimizing gifted identification practices on a Jacob K. Javits-funded collaborative project. Additionally, she advocates for use of open science practices to promote openness, transparency, and reproducibility in research in gifted education. In 2021, she was a recipient of the NAGC Carolyn Callahan Doctoral Student Award for an outstanding degree of scholarship in gifted education.

Catherine A. Little, Ph.D., is a Professor in Giftedness, Creativity, and Talent Development in the Department of Educational Psychology at the University of Connecticut. Catherine completed her graduate work at William and Mary, where she also served as Curriculum Coordinator for the Center for Gifted Education. Her research interests include professional learning, differentiation of curriculum and instruction for advanced learners, and classroom questioning practices. She co-edited the text *Content-Based Curriculum for High-Ability Learners* with Joyce VanTassel-Baska. Recently, she has been the Project Director for Project SPARK and Project LIFT, two Javits-funded initiatives focused on working with schools and teachers to recognize and respond to advanced academic potential in the early grades, particularly in students from underserved populations.

Tonya R. Moon, Ph.D., is Professor of Education in the School of Education and Human Development at the University of Virginia and Director of the Institutes on Academic Diversity. Tonya spends her professional career actively engaged in teaching assessment and research courses at the University and conducting research in K-12 classrooms investigating teachers' assessment practices and use of data for designing instructional plans. Tonya has co-led several multi-year national projects focusing on differentiating instruction with diverse student populations. She has published and presented widely on the topics of assessment, differentiation, and use of data for decision-making. Tonya works both within the US and internationally with educators on issues associated with assessment.

Rebecca L. O'Brien received her Ph.D. in Educational Psychology with a concentration in Giftedness, Creativity, and Talent Development from the University of Connecticut. She is an Assistant Professor of Graduate Studies in Education department, the Director for the Center for Gifted Education, and the Program Director for Gifted Education and Equitable Instruction at Whitworth University in Spokane, Washington. Her work primarily focuses on identifying and encouraging high potential in students from diverse backgrounds, improving teacher preparation, aligning instruction to assessment, and supporting in-service teacher learning and instruction. Prior to becoming a professor, she served as a math, science, and robotics teacher in an urban middle school.

Paula Olszewski-Kubilius, Ph.D., is the Director of the Center for Talent Development at Northwestern University and a professor in the School of Education and Social Policy. Over the past 35 years, she has worked to create outside-of-school program models to meet the need of diverse gifted learners. She has published extensively about issues in gifted education, with a particular focus on talent development for under-served gifted students. She has served as the editor of *Gifted Child Quarterly*, co-editor of the *Journal of Secondary Gifted Education* and on the editorial boards of *Gifted and Talented International*, *Roeper Review*, *Journal for the Education of the Gifted*, and *Gifted Child Today*. She currently is on the Board of Trustees of the Illinois Mathematics and Science Academy and is Past-President of the Illinois Association for the Gifted. She is also Past-President of the National Association for Gifted Children and received the Distinguished Scholar Award in 2009 from NAGC.

Scott J. Peters, Ph.D. is a Professor of Assessment and Research Methodology at the University of Wisconsin–Whitewater where he teaches courses on educational assessment, research methods, and gifted and talented education. His research work focuses on assessment and data use, gifted and talented student identification, equity within advanced learning opportunities, and educational policy. His scholarly work has appeared in the *Australian Educational Researcher*, *AERA Open*, *Teaching for High Potential*, the *British Journal of Educational Psychology*, *Gifted Child Quarterly*, the *Journal of Advanced Academics*, *Gifted and Talented International*, the *Journal of Career and Technical Education Research*, *Ed Leadership*, *Phi Delta Kappan*, *Gifted Child Today*, and *Pedagogies*. He is the first author of *Beyond Gifted Education: Designing and Implementing Advanced Academic Programs* and *Designing Gifted Education Programs and Services: From Purpose to Implementation*, both from Prufrock Press, and the co-author (with Jonathan Plucker) of *Excellence Gaps in Education: Expanding Opportunities for Talented Students*, published by Harvard Education Press.

Anne N. Rinn, Ph.D., is a Professor of Educational Psychology and Gifted Education at the University of North Texas, where she also serves as Director of the Office for Giftedness, Talent Development, and Creativity. She has an undergraduate degree in psychology from the University of Houston and a doctorate in educational psychology from Indiana University. She has authored many publications related to the social and emotional development of gifted and talented individuals and the psychosocial skills necessary for the development of talent. She is the author of *Social, Emotional, and Psychosocial Development of Gifted and Talented Individuals* (Rinn, 2020) and co-editor of *From Giftedness to Gifted Education: Reflecting Theory in Practice* (Plucker et al., 2017). She is also co-editor of the *Journal of Advanced Academics*.

Jennifer H. Robins, Ph.D., is the Director of the Center for Gifted Education and Talent Development and Clinical Assistant Professor in the Department of Educational Psychology at Baylor University. She received her doctorate in educational psychology with an emphasis in gifted education from Baylor University. Previously, she taught gifted elementary students, worked as Senior Editor at Prufrock Press, and was the Director of Publications and Professional Development at the Center for Gifted Education at William & Mary. Her textbook, *Introduction to Gifted Education*, which was coedited with Dr. Julia Roberts and Dr. Tracy Ford Inman, won the 2018 Texas Association for the Gifted and Talented Legacy Book Award in the scholar category. She is currently program chair for AERA's Research on Giftedness, Creativity, and Talent Development SIG and serves as Co-Editor-in-Chief with Dr. Jennifer L. Jolly of the *Journal for the Education of the Gifted*.

Gail R. Ryser, Ph.D., is a Research Scientist at Texas State University in the Methodology, Measurement, and Data Analysis unit located in the Office of Research and Sponsored Programs. This university-wide role involves assisting principal investigators in conceptualizing and writing the analytic portions of completive grant proposals and assisting faculty with designing, analyzing, and publishing their own research. Previously, she served as the Director of the Testing, Research-Support, and Evaluation at Texas State University. Dr. Ryser is the first author of the Scales for Identifying Gifted Students-Second Edition and the Test of Mathematical Abilities for Gifted Students-Second Edition.

Monica Simonds, M.Ed., is Director of Advanced Learning Programs and Services in Richardson ISD located north of Dallas. After receiving her Master's in Gifted Education, Monica taught gifted children in all grades and provided professional learning opportunities for the districts. She is a member of the National Association for Gifted Children and a board member for the Texas Association of Gifted and Talented.

She presents on gifted education topics such as leadership, equity, change, program design, and instructional design. She has published a book and contributed to a chapter in another as well as multiple articles. Monica is the 2017 recipient of the NAGC Gifted Coordinator award and the 2019 TAGT Administrator of the Gifted award.

Celeste D.C. Sodergren, M.A., is a researcher-practitioner who believes in utilizing the best and most current research to inform day-to-day practice in educational organizations. Currently, she is pursuing her doctoral degree in educational psychology, specializing in gifted and talented research, at Baylor University and serves as the Director of Advanced Academics in Waco ISD. Research interests include the influence of self-beliefs on teacher success and student academic performance, and leadership development in gifted and talented coordinators.

Rena F. Subotnik, Ph.D., is Director of the Center for Psychology in Schools and Education at the American Psychological Association (APA). One of the Center's missions is to generate public awareness, advocacy, clinical applications, and cutting-edge research ideas that enhance the achievement and performance of children and adolescents with gifts and talents in all domains. She has been supported in this work by the National Science Foundation, the American Psychological Foundation, the Association for Psychological Science, the Camille and Henry Dreyfus Foundation, and the Jack Kent Cooke Foundation. She is a Fellow of AERA, two Divisions of APA, and a Distinguished Scholar of the National Association for Gifted Children. Her work (with Paula Olszewski-Kubilius and Frank Worrell) is published in *Scientific American*, *Scientific American Mind*, *Annals of the New York Academy of Sciences*, *Frontiers in Psychology*, *Psychological Science in the Public Interest*, and the *Annual Review of Psychology.*

Tracey N. Sulak, Ph.D., is Clinical Associate Professor of Educational Psychology at Baylor University. She received a Master of Science in Curriculum and Instruction with a specialty in Gifted and Talented from the University of North Texas and a doctorate in Educational Psychology from Baylor University. She currently teaches courses on assessment and learning for practitioners and researchers and her research interest include talent development and broadening participation in STEM education as well as the role of environment in educational outcomes. Her work has been published in top tier journals ranging from the *Journal of Attention Disorders* to the *Journal of Family Issues*. Prior to joining the university, Tracey taught in public and private schools in classrooms serving preschool through eighth grade students; she taught students identified as special education or twice exceptional.

April Walker, Ph.D., is an adjunct instructor at Grand Canyon University in Curriculum and Instruction. She has 12 years of teaching experience in K-12 education. Much of her teaching experience focused on working with gifted students and encouraging critical thinking and self-assessment through inquiry, projects, and oral communication skills. She completed her doctoral degree at the University of North Texas in Curriculum and Instruction with a focus in Language and Literacy in 2020. Her primary research interests include critical thinking, advanced readers, adolescent literacy, and the nature of disciplinary literacy in history and social studies.

Frank C. Worrell, Ph.D., is a Professor in the Graduate School of Education where he also holds an affiliate appointment in the Social and Personality Area in the Department of Psychology. A licensed psychologist and certified school psychologist, his areas of interest include cultural identities, gifted education/talent development, scale development and validation, teacher effectiveness, time perspective, and the translation of psychological research findings into practice, and his scholarship crosses several subdisciplines of psychology, including cultural, developmental, educational, international, quantitative, school, and social and personality. Author of more than 200 scholarly works, Dr. Worrell is a Fellow of the American Educational Research Association, the American Psychological Association, and the Association for Psychological Science, and a member of the National Academy of Education and the Society for the Study of School Psychology. A former editor of *Review of Educational Research*, Dr. Worrell is also the 2021 President-elect of the American Psychological Association.

CHAPTER

1

Using Assessment as a Framework for Standards-Based Program Development in Gifted Education

Susan K. Johnsen and Joyce VanTassel-Baska

Importance of Assessments

The first area of assessment that is critical to the field is that of identification—who are the gifted? The problem of **identification** has haunted the field since its nascent grass roots emergence in the 1970s. Problems related to the inclusion of underrepresented groups, the need for early identification, the need for matching or aligning identification to program options, and the need to examine specific domain abilities for inclusion in programs all gathered force during this period. There were later cries even for abandoning identification altogether (see Borland, 1985).

Yet identification has still retained its prominence as one of the first major steps in program development (VanTassel-Baska & Baska, 2020) for several reasons. One of these is its power in finding the most promising students to be served in different domains. With the advent of Talent Searches at major universities nationwide, the field could now efficiently find high achieving students who represented a continuum of ability in verbal and mathematical areas who scored at the 95–99th percentile on grade level achievement measures but who also demonstrated advanced cognitive ability in these subject areas 2–6 years beyond their placement. Thus the use of achievement tools in concert with off-level aptitude measures became popular for finding students who were gifted in those areas. Because the program was voluntary and external to the school district, it never gained ground as a universal approach to identification even though 100,000 seventh graders are still assessed each year and provided programming through multiple options both online and in person.

In addition to the power of identification for finding top learners in math and English language arts, systems have also been developed that prove to be useful in finding students strong in the cluster areas of science and spatial areas (see Lubinski, 2016) leading to programs in robotics, LEGOs, and STEM that address these types of abilities more directly. Arts abilities are assessed through multiple measures, including a

DOI: 10.4324/9781003285991-1

portfolio of products and performances and teacher recommendation. Less common have been protocols for identifying leadership as a talent area separate from advanced content indicators. More routinely, school districts have selected the content areas for learning within which they wish to serve the gifted—typically ELA and math (Siegle et al., 2016).

Finally, identification has been an important tool in designing programs that match the specialized abilities of students found in a particular locale. A school district that has a cluster of math whizzes at 4^{th} grade clearly needs a set of curriculum experiences planned along the continuum from 4^{th} grade through high school to ensure their optimal educational opportunity. Knowing who and where they are as well as how advanced they are has become important planning data for program developers. Thus identification has become the trigger for further program development across the sequence of schooling years.

Over the past 40 years**, learning assessments** have assumed greater importance at the national, state, and local levels. Since the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), states have demanded more rigorous expectations and measurable benchmarks for student achievement. The passage of the No Child Left Behind Act and the subsequent Every Student Succeeds Act (ESSA) have required students to be taught to high academic standards that prepare them to succeed in college and careers and have mandated statewide assessments that measure students' progress toward those high standards (U.S. Department of Education, 2015). Moreover, ESSA requires states to submit plans that improve the skills of teachers, principals, or other school leaders to identify students with specific learning needs (e.g., children with disabilities, English learners, students who are gifted and talented, and students with low literacy levels) and provide instruction based on the needs of these students (U.S. Department of Education, 2017). In their plans, some of the states have described methods for meeting the specific needs of gifted students that incorporate differentiated assessments. For example, the Kansas Department of Education (2018) uses a Multi-Tier System of Supports (MTSS) framework to identify and address the needs of all students, including those who are gifted and talented. Using the MTSS framework, educators use data to monitor academic, behavioral, and social emotional learning needs and make decisions that inform instructional decision making. The Illinois Department of Education (2017) also emphasizes supporting the social and emotional development as well as differentiated instruction. Similarly, the South Dakota Department of Education (2018) supports schools in implementing individualized education opportunities described as Mass Customized Learning. Using formative assessments, teachers are encouraged to allow students to work at their own pace to address their academic needs and interests. Under ESSA, all states are required to submit plans to the federal government, showing how they will assess students' progress, with most states explicitly addressing how educators should be supported in providing effective, differentiated instruction to gifted learners (Kaul & Davis, 2018). States and local districts now have opportunities to develop and implement assessments that identify and examine the learning progress of gifted students.

The last area of assessment that has gained prominence as programs have developed and grown is **program evaluation**. How do we know that the program is doing what it was designed to do? How do we assess stakeholder satisfaction with the program? How do we know that all parts of the program (e.g., identification, instruction) are working effectively? These questions can only be answered through an annual assessment of the operational gifted program. Thus the importance of evaluation as a part of the program development mechanism has been shown (VanTassel-Baska & Hubbard, 2019). Its importance is at least three-fold in providing the following:

- A report card to school districts internally on the progress made by both students and the program itself in the area of advanced learning so that changes may be made for improvement.
- An accountability to the community on important services provided to top students in the district; an ongoing record of accomplishment of students and schools in different realms of excellence is provided, trophies and awards that testify to the school's prowess in learning provisions.
- A tool for judging how future funds might best be allocated for program stability or expansion.

All of these reasons provide an important purpose for regular evaluations of gifted programs to be conducted. Improvement of gifted programs is not possible in the absence of data that provide educators with the knowledge of current practice. Best practice may be assessed by using the national standards developed for the field and analyzing any given program against those standards.

Assessment Standards in Gifted Education

The NAGC Pre-K to Grade 12 Programming standards (NAGC, 2019a) offer guidance in all three areas of assessment that this book examines: identification, learning progress, and evaluation. (See Appendix A for an alignment between the standards.) The standards are organized into six areas: Learning and Development, Assessment, Curriculum Planning and Instruction, Learning Environments, Programming, and Professional Learning. Within each of these areas, evidence-based practices have been identified and linked to specific student outcomes. In this way, educators can determine the effectiveness of the practice by assessing students' growth in academic, social and emotional, and psychosocial areas.

Identification

Identification is the process of finding students who have needs for or would benefit from advanced programming or services to develop their gifts and talents (NAGC, 2019b). In the area of identification, the first student outcome focuses on the student's access to the identification process so that each campus is proportionally represented (NAGC, 2.1, p. 8). Access is influenced by these evidence-based practices: (a) educators' instructional activities that encourage students to express characteristics and behaviors associated with giftedness; (b) parents/guardians understanding of gifted characteristics and the need for programming; and (c) the use of universal screening and multiple indicators across multiple entry points. Research suggests that when teachers differentiate, students with gifts and talents are more likely to show themselves and be recognized (Bianco & Harris, 2014; Grantham, 2003; Hertzog, 2005) Teachers also need to know that giftedness assumes many forms and not stereotype the characteristics of a "gifted child" (e.g., academic achievers, higher socioeconomic groups, verbal and well behaved, similar cultural backgrounds; Grissom & Redding, 2016; Kitano, 2003; Peters & Engerrand, 2016; Speirs Neumeister et al., 2007; Swanson, 2006; Yoon & Gentry, 2009).

To ensure that students are able to demonstrate their talents at home, parents/guardians also need to be educated about diverse characteristics and behaviors of children with gifts and talents, how to nurture their children, and the importance of gifted education (Jolly & Matthews, 2012; Siegle et al., 2016; Wilder, 2014). Parents who have fewer financial resources may need assistance in finding after school and summer enrichment programs that offer scholarships (Johnsen et al., 2007). In addition to professional learning and parent education, schools need to adopt identification practices that use universal screening and multiple indicators of potential at multiple entry points to ensure that all students have equal access (see Standard 2.1.3). Universal screening or consideration is where one or more assessments are administered to *all* of the students as separate assessments, as part of a Response to Intervention process, or as part of a talent development program (Card & Giuliano, 2015; Coleman & Shah-Coltrane, 2010; Sulak, 2014). These assessments are intended to bypass referral systems where teachers or parents may overlook gifted and talented students who do not fit traditional stereotypes.

The second student outcome (NAGC, 2.2, p. 8) focuses on the use of assessments that identify student's interests, strengths, and needs. Eight evidence-based practices encourage educators to (a) develop comprehensive, cohesive, and ongoing policies and procedures; (b) select assessments that relate to services and student characteristics; (c) include qualitative and quantitative information from a variety of sources; (d) use assessments related to above-level performance; (e) minimize bias; (f) understand exceptionalities and use dynamic assessments; (g) interpret multiple assessments in different domains; (h) inform parents/guardians about the identification process and elicit evidence. Comprehensive and cohesive procedures require that educators align assessments across grade levels and provide programming options in all domains. Once these procedures are developed, educators need to develop policies and procedures that describe the process such as consent, committee reviews, retention, reassessment, exiting from program services, and appeals for entering and exiting gifted program services. Assessments should be selected based on the characteristics of the students and available services. For example, if the majority of students are from diverse backgrounds (e. g., English language learners, low income, minorities), then alternative types of instruments might need to be considered such as performance assessments or those that are nonverbal or linguistically-reduced. In addition, multiple sources of information (e.g., parent/guardian, teachers, peers) and different types of information (e.g., qualitative and quantitative) provide a broader description of the student and a more comprehensive view of the student's behaviors across settings (Coleman & Cross, 2005; Johnsen, 2018; Ryser, 2018b). Since most tests are intended for determining growth at a particular grade level, multi-grade level and off-level tests provide more difficult items that are able to discriminate better among students with gifts and talents (Olszewski-Kubilius & Kulieke, 2008). All assessments used during the identification process need to minimize bias, increase equity, and provide information describing their technical adequacy—validity, reliability, and norming (Ryser, 2018a). Another important assessment tool is the use of dynamic assessments where the teacher focuses on the interaction between the student and the task to understand the student's academic strengths and needs (Swanson & Lussier, 2001). These are particularly helpful when identifying students from underrepresented populations (Lidz & Macrine, 2001). After collecting the assessments, educators need to know how to interpret all of the information to determine the best programming and services for each student (Johnsen, 2018). Finally, since parents offer perspectives about their child's interests and potential, they need to be actively involved in gathering information (Jolly & Matthews, 2014; Lee & Olszewski-Kubilius, 2006).

The third Student Outcome in the area of identification examines the representation of students from diverse groups in the school's program. Evidence-based practices focus on policies designed to foster equity and approaches and tools that minimize bias. These approaches include front-loading talent development activities or providing emergent talent development experiences (Briggs et al., 2008; Olszewski-Kubilius & Clarenbach, 2012; Subotnik et al., 2011; Siegle et al., 2016); universal screening (Card & Giuliano; 2015; McBee, 2016); locally developed norms (Peters & Engerrand, 2016); the use of assessment tools in the child's preferred language or in nonverbal or performance formats (Ryser, 2018a; VanTassel-Baska et al., 2007; VanTassel-Baska et al., 2002); and building relationships with students and their parents/guardians to understand their strengths, needs, and interests (den Brok et al., 2002; Ford & Trotman, 2001).

Learning Progress

In addition to identification, assessments are used to examine the students' learning progress or growth in their domain of talent and/or area(s) of interest. Twelve of the student outcomes explicitly examine the assessment of students' progress in cognitive, social-emotional, psychosocial (e.g., personal and social competence), and career development areas. These student outcomes result from multiple and ongoing assessments; self-assessment; meaningful and challenging learning activities that address students' unique characteristics and needs; high-quality curricular resources related to the students' goals; comprehensive, cohesive, and coordinated programming and services; and well prepared educators. In examining the 40 evidence-based practices influencing these outcomes, the standards included multiple forms and uses of assessments for measuring learner progress:

- *Qualitative and quantitative assessments*. Educators interpret qualitative and quantitative assessments to develop a student profile of interests, strengths and needs to plan individual interventions. These various types of assessments are from multiple sources and contexts (Coxbill et al., 2013; Johnsen, 2018; Olszewski-Kubilius & Kulieke, 2008; VanTassel-Baska et al., 2007). A variety of instruments are useful when assessing learning progress. Traditional assessments, such as standardized tests, are effective for measuring reasoning skills and content-area achievement; qualitative methods, such as performance based assessments and tests of critical thinking, are needed to measure more complex thinking (Jolly & Kettler, 2004; Kim et al., 2014).
- *Pre-, formative, and summative assessments*. Educators use differentiated pre-assessments, formative assessments, and summative assessments to identify student's strengths and needs, develop differentiated content and learning experiences, and adjust instructional plans based on progress monitoring (Inman & Roberts, 2016; McCoach et al., 2013; Tieso, 2005; Tomlinson et al., 2003). Preassessment is important in finding the student's level and moving forward with a curriculum tailored to the student's strengths and needs. The differentiated content and learning experiences include acceleration, enrichment, grouping, individualized learning, independent research, mentorships, and digital learning options that develop students' talent areas (Renzulli & Reis, 2003; Siegle & McCoach, 2005; Steenbergen-Hu et al., 2016).
- *Product- and performance-based assessments*. Educators use differentiated ongoing product-based, performance-based, and above level standardized assessments to measure progress (Kim et al., 2014; VanTassel-Baska et al., 2002; VanTassel-Baska et al., 2007). These assessments include off-level standardized achievement measures;

end-of-course or AP exams; rubrics for assessing complex products and performance; critical or creative-thinking measures to assess process skills; pre- and post-assessments, portfolio assessments, or student self-assessments such as journals, written products, or surveys to examine students' performances over time. Advocates for these assessments suggest that students represent learning in multiple and creative ways and are needed to assess student growth (Duggan, 2007; Ryser & Rambo-Hernandez, 2014). The assessments need to be related to the learning outcome and include advanced, higher-level thinking and open-ended problem solving tasks that challenge gifted learners to demonstrate deep meaningful learning (Duggan, 2007; Kaplan, 2008; VanTassel-Baska, 2014).

- *Self-assessments.* Students identify their interests, strengths, and needs and set personal goals, keep records, and monitor their own learning progress (Johnsen, 2009; Simonton, 2000; van Deur, 2011; VanTassel-Baska, 2009). Newman (2004) found that when students were involved in self-assessment during the creative process their products were of higher quality. Likewise, Sriraman (2004) discovered that when mathematically gifted students were asked to reflect upon and analyze their own thinking processes, they were able to produce at a level characteristic of professional mathematicians. Being given the opportunity to create both assignments and assessments for their own products has also proven motivating to gifted learners (Thompson & McDonald, 2007). The most creative and expressive products resulted from student-constructed assignments.
- *Career assessments.* Educators implement learning progressions in specific domains of talent, provide students with college and career guidance, and collaborate with students to identify college and career goals (Greene, 2003; Jung, 2014; Maxwell, 2007; Muratori & Smith, 2015). Information from these assessments is used to differentiate counseling, enhance personal growth and development, establish mentorships, and channel students into promising career trajectories.
- *Modifications and accommodations of assessments.* The practices also address issues related to diversity that would affect assessment such as accommodations and modifications for learning differences, underachieving students, twice exceptional students, English language learners, highly gifted students, and students from lower socioeconomic and different cultural and racial backgrounds (Bianco & Harris, 2014; Coleman & Hughes, 2009; Ford, 2006; Hertzog, 2005). Accommodations do not change the assessment (e.g., extended time) whereas modifications may change the assessment itself (e.g., answering different test questions).
- *Interpretation of assessments.* Educators communicate assessment information to parents/guardians and other professionals and collaborate to develop a continuum of programming and services (Boazman, 2015; Campbell & Verna, 2007; Jolly & Matthews, 2012). Assessment interpretation and explanations are guided by defensible and equitable principles and practices that nurture each student's advanced abilities and needs.

Evaluation

Program evaluation is the assessment of a system to ensure its components are aligned to the standards, fully implemented, and have the desired effects. Within the Programming Standards, two student outcomes address evaluation: (a) students have access to advanced academic services and (b) demonstrate yearly progress as a result of high-quality programming and services, ongoing evaluation, and program improvement. The five evidence-based practices to ensure these outcomes focus on the fidelity of implementation, disaggregating yearly progress data to examine differences in groups and services, reliable evaluation instruments, resources for conducting the evaluation, the expertise of the evaluators, dissemination and interpretation, and using the results of the evaluation. Examples of quality evaluations, which incorporate some or all of these practices, may be found in the literature: evaluations of curriculum (VanTassel-Baska et al., 2000), mathematics (STEM) school programs (Subotnik et al., 2016), online programs (Sanderson & Greenberger, 2011), fidelity of implementation (Missett & Foster, 2015), and gifted programming as a whole (Reis & Callahan, 2004; Robinson et al., 2014). Evaluations need to include their purpose—summative or formative (Gallagher, 2006); use instruments that are reliable and valid for the purpose (Speirs Neumeister & Burney, 2019; VanTassel-Baska et al., 2006; Warne, 2014); collect information that encompasses the key program and curricular components (Walker & Vander Ploeg, 2015); consider the unique characteristics of the population being served (Briggs et al., 2008; Robinson et al., 2014); and measure progress of outcomes toward the program's goals (Robinson et al., 2014; Subotnik et al., 2016). Since treatment fidelity is key in making decisions, evaluations need to be conducted by individuals who are experts in gifted and talented education (Gallagher, 2006; Kettler, 2016; Missett & Foster, 2015). A recent survey conducted by the National Association for Gifted Children's Professional Standards Committee suggests that one of the top three uses of the Programming standards is for program evaluation (Corwith & Johnsen, 2021). They provide a structure for defining critical benchmarks; developing policies, rules, and procedures; identifying and implementing research-based practices with fidelity; and providing opportunities for collaboration and program improvement (Corwith & Johnsen, 2021). Matthews and Shaunessy (2010) caution that while standards are necessary to develop and improve programs, they need to be user-friendly for practitioners to translate and implement.

An Overview of the Chapters and Their Relationship to the Standards

Assessments related to identification are addressed in the first six chapters. Preceding this section, we provide criteria for an identification system: preparation and pre-identification, screening, interpretation of assessments, and program services.

Following this introduction, Lee and Peters describe two approaches in Chapter 2 to broadening the pool of students considered for gifted programs—universal screening and universal consideration. They suggest that both approaches are more equitable than traditional two-phase identification systems relying on teacher or parent referral. They present how universal screening procedures can be designed and how to maximize effectiveness, minimize cost, and improve equity of the identified population.

In examining quantitative assessments in the third chapter, Lakin, Simonds, and Caviness-French propose that gifted education coordinators need a variety of data to understand the students' current academic performance and their potential for advanced academic performance. They provide an overview of a few broad categories of assessments, including measures of ability and achievement. They emphasize that student characteristics should drive the choice of an assessment and the types of information needed to inform service and program decisions. To make data-based and equity-informed decisions, coordinators need to be informed about what is being measured and how scores are influenced by student background and prior educational experiences.

The fourth chapter describes qualitative assessments that are also used in the identification process. Little, O'Brien, and Kearney address how educators may use evidence from curriculum-based performance tasks and student engagement with classroom learning activities as part of an identification system. They provide guidance to teachers about how they might document and interpret evidence from student performance and use these types of assessment as part of an overall identification system.

In Chapter 5, Ryser examines the usefulness of using rating scales as one measure to identify students as needing gifted and talented programming. She suggests that an advantage of using rating scales is that they allow educators to combine many observations of students' behavior in an efficient way. Information derived from rating scales complement achievement or aptitude measures because the information is based on observations in the school or home setting, which makes it possible to identify talent from multiple perspectives. She also discusses a method for combining and interpreting achievement or aptitude measures and rating scales.

In the sixth chapter, we examine challenges and possible solutions for including more students from underrepresented populations in programs for gifted students. Along with research, we use interviews with school directors and coordinators from rural/urban settings to provide a framework for examining local identification protocols, challenges, and possible solutions. Respondents share contextual issues that have made progress on this issue difficult. Some of these issues relate to policy and rules, leadership, tools and data sources, the match between student characteristics and assessment instruments, the interpretation of information, and legal challenges.

At the conclusion of this section, we incorporate the major concepts presented in the chapters. We respond to the introductory criteria by providing a figure showing the nature of the identification cycle and describing specific steps for practitioners to consider. We emphasize the importance of developing programs concurrently with identification. If program options are already in place, then they must be reviewed and based on the students' data to ensure that they have been placed in relevant classes.

Learning progress is examined in Chapters 7–13. We introduce this section by defining learning progressions and providing assumptions held about gifted programs. We emphasize the importance of differentiated assessments being accompanied by differentiated curriculum and instructional practices. Along with research, specific models are shared to ensure that school districts have a consistent approach to providing advanced learning.

In Chapter 7, Kettler and Lamb identify how educators can use domain-specific assessments of creativity as part of gifted education programs to recognize and develop talent. The authors describe three ways of measuring domain-specific creativity consistent with creativity research traditions; creative person, creative process, and creative products. They define and describe (a) standards-based assessment of content standards infused with creativity standards in core domains and the arts; (b) consensual assessment techniques that can be applied to authentic assessments; (c) domain-specific rubrics to measure domain-specific creativity; (d) creative-strengths profiles to articulate student's creative ability domains and match strengths to program services; (e) students' self-assessments and (f) open-ended standardized assessments.

Curriculum-based assessments (CBA) are examined in Chapter 8. Sulak, Robins, and Hebda describe this class of assessments as measuring skills and knowledge related to the curriculum. CBAs provide teachers with important, specific information about their students' background, knowledge, and skills, allowing them to adapt to meet each student's needs at different tiers of service. Practical implications for CBAs include identifying specific areas of talent, identifying an instructional level that matches a student's needs for acceleration or curricular compacting, and measuring progress on an independent project. Although many types of CBAs are proposed, they should be appropriately challenging, intentionally designed to inform instruction, and include multiple measures of qualitative and quantitative data. The authors conclude with specific examples of performance-based assessments in core subject areas.

In Chapter 9, Brown focuses on the rationale for using advanced assessments for gifted students at the secondary level and provides examples of traditional curriculum-based assessments, above grade level content-based assessments, problem-based assessments, portfolio-based assessments, and the role of competitions. She discusses how assessment information might inform instruction and promote advanced learning across curricula domains.

Moon, Brighton, and Hock describe project-based tasks and rubrics in Chapter 10. They include performance assessments that offer opportunities for students to investigate issues, problems or opportunities embedded within contexts that have relevance to their interests and are appropriate for their age. Potential benefits for student learning include: (a) demonstrating developing expertise of a discipline; (b) increasing student motivation in both the learning and assessment contexts; (c) developing students' critical, creative, and metacognitive skills, and (d) offering opportunities for increased access to post-secondary careers. Despite these benefits, the authors also identify potential liabilities such as issues of equity and access. They conclude by providing examples for employing these tools within schools and programs.

In assessing the learning progress of students with special needs, Farah introduces important concepts that should be considered when developing, administering, and evaluating assessments in Chapter 11. She discusses the definition, purpose, and characteristics of assessment accommodation and modification in terms of equitable approaches that minimize bias and provide learning progress opportunities for all students. She provides research findings and practical examples that relate to how accommodations and modifications might be used to provide for a wide range of learners so they might access appropriate assessment opportunities to demonstrate learning progress. She concludes with recommendations for coordinators and directors of gifted education programs.

Rinn describes student self-assessment in Chapter 12. She includes a variety of techniques and tools (e.g., rubrics, learning contracts) through which students assess and evaluate their own learning processes and products. She stresses the importance of training students to use self-assessment, so they not only experience an increase in learning and academic performance, but also an increase in self-regulated learning strategies, self-efficacy, and agency/empowerment. In this chapter, she provides practical examples and tips for effectively using student self-assessment in the classroom and recommendations for coordinators and directors of gifted education and advanced academic programs.

In Chapter 13, Boswell, Adams, and Christopher examine the use and value of state assessments of learning. They describe how state assessments for learning vary as widely as the states themselves and their approaches to gifted programming and services. The value of these assessments differs according to policies and practices and their fidelity in being implemented at the school level. The authors describe criteria for judging each state's assessments in terms of their efficacy and utility in gifted education and their ability to address equity issues. They present examples of state assessments that meet their criteria. They conclude by exploring challenges that each state faces related to implementation and provide recommendations to facilitate the work of coordinators and directors of gifted education.

In Chapter 14, Olszewski-Kubilius, Subotnik, and Worrell discuss the assessment of talent development trajectories. The trajectory requires taking talent from the potential for high achievement to developed competencies, expertise, and adult creative productivity. The authors describe special inputs and supports that include within and outside of school programs, lessons, teachers with domain expertise, and mentoring and guidance all along the talent development trajectory. They examine this important question: How do we know that individuals are making sufficient progress at each stage of talent development, enabling them to transition to higher stages? In this chapter, they provide criteria and tools that teachers can use to assess if their students are acquiring the domain knowledge and psychosocial skills needed to stay on domain- specific talent development trajectories.

Similar to the previous section's conclusion, we incorporate the major concepts presented in the chapters. We then provide a figure showing the learning progress cycle in gifted programs and services and emphasize the importance of selecting assessment tools that meet technical adequacy standards.

In introducing the final section on evaluation, we include the criteria for developing and implementing an evaluation: purpose, evaluation questions, data sources, assessment instruments, expected outcomes, timeline (task, person responsible, dates), results, and action plan.

Following this introduction, in Chapter 15 Adams begins with the purposes and criteria for evaluating tools to improve programs using the 2019 NAGC PreK-Grade 12 Standards for Gifted Programming. She describes (a) different types and characteristics of evaluation tools at the local and state levels, (c) how each meet the criteria, and (c) how they might be interpreted and used effectively to improve programs and services, particularly for diverse learners. Types and samples of evaluation tools are included. Adams also includes examples of successful tools in improving programming for all gifted learners in general, and gifted learners from diverse backgrounds in particular.

In Chapter 16, Callahan illustrates how school districts have used assessments to improve programming and services. She begins with examples of assessments used across a variety of decision-making situations. Her examples include (a) student data generated from identifying gifted students, modifying instruction, modifying delivery of instruction/modifying program structure, and evaluating outcomes; and (b) teacher data for modifying professional development and/or provision of resources. Callahan concludes with lessons that emphasize the importance of planning and communicating the process of using data for program change.

Similar to other conclusions we incorporate the major concepts presented in the chapters that address the introduction criteria and examine emergent issues from gifted program evaluations.

In the concluding chapter, we review the key insights from the authors in each section of the book on the value and use of assessments aligned with national best practice via the standards and reported research. We also summarize the major principles, practical implications that emerge from these principles, and challenges identified by the authors. The different types of assessments are then interwoven into a practical model that delineates how assessment work together for developing and improving gifted education programming.

Summary and Conclusion

This handbook on assessment contains important benchmarks for districts to use in designing, developing, and implementing their systems of identification, learning progress, and program evaluations. It provides a roadmap for ensuring alignment to the NAGC standards related to these three aspects of assessment, and its chapters provide an important discussion of the major issues that grip the field of gifted education as it works to carry out effective program development practices overall. While the 2006 version of the book was meant to examine multiple approaches to nontraditional identification and learning progress, it focused more on specific tools for use rather than on how assessment is an integrated process that considers the students at the beginning, middle, and end of their educational experiences.

This book also fills a gap in addressing the breadth and depth of what assessment of gifted learners and their programs and services mean. It provides a clear picture of all three types of assessment as part of one comprehensive program development framework, focused on the characteristics and needs of a special population requiring a differentiated program of study that sustains them through their years in school and beyond into career paths and trajectories well-matched to their aptitudes and interests. It suggests a mechanism for ensuring annual progress and multi-year growth in learning areas and improvement in programmatic processes. Many resources address identification but do not embed principles that provide an optimal match between learning assessments to program opportunities that children may need at a given stage of development to measure their advanced learning progress. Moreover, resources are available that may offer guidelines for program evaluation as a separate enterprise rather than as an integral part of the overall process of assessment. This book offers a totally integrated model for examining who should be identified for gifted programs, what their programs should look like, how we know they are learning from them, and how the program is advancing on issues like underrepresentation and equity of opportunities.

Appendix A

Alignment of Chapters to Important Elements of the 2019 NAGC Programming Standards

To determine whether or not these chapters addressed the important elements of the 2019 NAGC standards that incorporated assessment, we reviewed each of the gifted student outcomes and evidence-based practices to examine the relationships. While every student outcome would incorporate some form of assessment, we included the outcome if one or more of the evidence-based practices focused on some form of assessment (see Table 1.1). In reviewing Table 1.1, all of the chapters are aligned to the standards with some assessments included more frequently (Chapter 10. Project-Based Learning Tasks and Rubric) than others (Chapter 5. Rating Scales). This variation is most likely due to the emphasis in different student outcomes and the generalizability of the assessments themselves.

TABLE 1.1 Alignment of chapters with 2019 NAGC programming standards

Gifted student outcomes	Evidence-based practice elements	Related chapter
1.1.Recognize their strengths, needs, interests	Students identify intellectual, academic, creative, leadership, artistic domains.	7, 9, 10, 12
1.2. Recognize influences on how they learn.	Influences include learning environment and culture-based learning needs.	12
1.4. Access outside resources	Resources include families, mentors, experts, programs	11
1.6. Identify future career goals and access resources.	Resources relate to learning progressions, career, and psychosocial skills development.	9, 10, 11
2.1. Have access to identification process.	Access includes universal screening and multiple assessment indicators.	2
2.2. Have assessments and services that match their interests, strengths, and needs.	Quantitative, qualitative, rating scales, above-level, and dynamic assessments that minimize bias and use multiple sources.	3, 4, 5, 6, 8, 14
2.3. Identified represent diverse backgrounds.	Approaches include policies to foster equity and front loading, universal screening, local norms, nonverbal assessments, and building relationships.	2, 3, 4, 6, 13
2.4. Demonstrate growth as measured by ongoing assessments.	Ongoing assessments include formative, product- and performance-based, standardized, above-level and adaptive, qualitative and quantitative.	3, 4, 14
2.5. Self-assess learning progress.	Self-assessment includes setting personal goals, keeping records, and monitoring progress.	10, 12
3.1. Demonstrate academic growth commensurate with abilities.	Use standards and assessments (pre-, formative summative) to adapt, modify, replace content to differentiate and monitor progress.	8, 9, 10, 12, 13, 14
3.2. Demonstrate growth in social and emotional and psychosocial skills.	Learning experiences are designed to develop psychosocial skills for each stage of learning development.	12
3.5. Become independent investigators.	Students use self-assessment and monitor their learning in these skills: metacognition, goal setting, independent research.	9, 10, 12
4.1. Demonstrate growth in personal competence and dispositions for exceptional academic and creative production.	Assessment feedback promotes perseverance and resilience with opportunities for the development of achievement identities (mentors/role models).	7, 10, 11, 12
4.5. Develop competence in interpersonal and technical communication, advanced oral and written skills, and creative expression.	Resources reflect diversity of the student population and include tools for expressing higher-level thinking and creative productivity.	7, 9, 10, 11, 14
5.1. Demonstrate growth as a result of comprehensive programming and services.	Services include acceleration, clusters, individualized learning, mentorships, online courses, independent study.	9, 10, 11

TABLE 1.1 Cont.

Gifted student outcomes	Evidence-based practice elements	Related chapter
5.2. Demonstrate progress as a result of cohesive and coordinated services.	Services respond to students' different levels with programs evaluated by all educators.	16
5.3. Create future career goals and identify talent pathways.	Mentorships, internships, and career programming are matched to students.	9, 10, 11
5.7. Demonstrate yearly progress resulting from programming and services.	Assessments are reliable and valid with progress data disaggregated and shared.	15, 16
5.8. Access services resulting from evaluation and program improvements.	Educators create, implement, and disseminate evaluation plans that are purposeful and evaluate implementation components. Two of the components include ongoing assessment of student learning and school equity.	6, 15, 16

References

Bianco, M., & Harris, B. (2014). Strength-based RtI: Developing gifted potential in Spanish-speaking English language learners. *Gifted Child Today*, *37*(3), 169–176. https://doi.org/10.1177/1076217514530115

Boazman, J. (2015). Important voices: Gifted children and parents share what they need. *Parenting for High Potential*, *4*(7), 2–3, 22–23.

Borland, J. H. (1997). The construct of giftedness. *Peabody Journal of Education*, *72*(3–4), 6–20. https://doi.org/10.1080/0161956X1997.9681863

Briggs, C. J., Reis, S. M., & Sullivan, E. E. (2008). A national view of promising programs and practices for culturally, linguistically, and ethnically diverse gifted and talented students. *Gifted Child Quarterly*, *52*(2), 131–145. https://doi.org/10.1177/0016986208316037

Campbell, J. R., & Verna, M. A. (2007). Effective parental influence: Academic home climate linked to children's achievement. *Educational Research and Evaluation*, *13*(6), 501–519. https://doi.org/10.1080/13803610701785949

Card, D., & Giuliano, L. (2015). *Can universal screening increase the representation of low income and minority students in gifted education?* (NBER Working Paper No. 21519). Cambridge, MA; National Bureau of Economic Research. www.nber.org/papers/w21519

Coleman, L. J., & Cross, T. L. (2005). *Being gifted in school: An introduction to development, guidance, and teaching*. Prufrock Press.

Coleman, M. R., & Hughes, C. E. (2009). Meeting the needs of gifted students within an RtI framework, *Gifted Child Today*, *32*(1), 14–17. https://doi.org/10.1177/107621750903200306

Coleman, M. R., & Shah-Coltrane, S. (2010). *U-STARS~PLUS professional development kit*. Council for Exceptional Children.

Corwith, S., & Johnsen, S. K. (2021). Gifted programming standards. In J. A. Plucker and C. M. Callahan (Eds.), *Critical issues and practices in gifted education* (3rd ed., pp. 229–247). Prufrock Press.

Coxbill, E., Chamberlin, S. A., & Weatherford, J. (2013). Using model-eliciting activities as a tool to identify and develop mathematically creative students. *Journal for the Education of the Gifted*, *36*(2), 176–197. https://doi.org/10.1177/0162353213480433

den Brok, P., Levy, J., Rodriguez, R. & Wubbels, T. (2002). Perceptions of Asian-American and Hispanic-American teachers and their students on teacher interpersonal communication style. *Teaching and Teacher Education*, *18*(4), 447–467. https://doi.org/10.1016/S0742-051X(02)00009-4

Duggan, T. J. (2007). Ways of knowing: Exploring artistic representation of concepts. *Gifted Child Today*, *30*(4), 56–63. https://doi.org/10.4219/gct-2007-486

Ford, D. Y., & Trotman, M. F. (2001). Teachers of gifted students: Suggested multicultural characteristics and competencies. *Roeper Review*, *23*(4), 235–239. https://doi.org/10.1080/02783190109554111

Ford, D. Y. (2006). Creating culturally responsive classrooms for gifted students. *Understanding Our Gifted*, *19*(1), 10–14. https://eric.ed.gov/?id=EJ846094

Gallagher, J. J. (2006). According to Jim Gallagher: How to shoot oneself in the foot with program evaluation. *Roeper Review*, *28*(3), 122–124. https://doi.org/10.1080/02783190609554350

Grantham, T. C. (2003). Increasing Black student enrollment in gifted programs: An exploration of the Pulaski County special school district's advocacy efforts. *Gifted Child Quarterly*, *47*(1), 46–65. https://doi.org/10.1177/001698620304700106

Greene, M. J. (2003). Gifted adrift? Career counseling for the gifted and talented. *Roeper Review*, *25*(2), 66–72. https://doi.org/10.1080/02783190309554201

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*(1), 1–25. https://doi.org/10.1177/2332858415622175

Hertzog, N. B. (2005). Equity and access: Creating general education classrooms responsive to potential giftedness. *Journal for the Education of the Gifted*, *29*(2), 213–257. https://doi.org/10.1177/016235320502900205

Illinois Board of Education. (2017, August 30). *Illinois Board of Education state template for the consolidated state plan under the Every Student Succeeds Act*. www2.ed.gov/admins/lead/account/stateplan17/ilconsolidatedstateplan.pdf

Inman, T. F., & Roberts, J. L. (2016). Authentic, formative, and informative assessment of advanced learning. In T. Kettler (Ed.), *Modern curriculum for gifted and advanced academic students* (pp. 205–235). Prufrock Press.

Johnsen, S. K. (Ed.). (2018). *Identifying gifted students: A practical guide* (3rd ed.). Prufrock Press.

Johnsen, S. K. (2009). Portfolio assessment of gifted students. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 227–257). Waco, TX: Prufrock Press.

Johnsen, S. K., Feuerbacher, S., & Witte, M. M. (2007). Increasing the retention of gifted students from low income backgrounds in a university program for the gifted: The UYP project. In J. VanTassel-Baska (Ed.), *Serving gifted learners beyond the traditional classroom: A guide to alternative programs and services* (pp. 55–79). Prufrock Press.

Jolly, J., & Kettler, T. (2004). Authentic assessment of leadership in problem-solving groups. *Gifted Child Today*, *27*(1), 32–39. https://doi.org/10.1177/107621750402700110

Jolly, J. L., & Matthews, M. S. (2012). A critique of the literature on parenting gifted learners. *Journal for the Education of the Gifted*, *35*(3), 259–290. https://doi.org/10.1177/0162353212451703

Jolly, J. L., & Matthews, M. S. (2014). Parenting. In J. A. Plucker, & C. M. Callahan (2014), *Critical issues and practices in gifted education: What the research says* (2nd ed., pp. 481–492). Prufrock Press.

Jung, J. Y. (2014). Modeling the occupational/career decision-making processes of intellectually gifted adolescents: A competing models strategy. *Journal for the Education of the Gifted*, *37*(2), 128–152. https://doi.org/10.1177/0162353214529045

Kansas Department of Education. (2018, January 8). *Kansas consolidated state plan: The elementary and secondary education act of 1965, as amended by the Every Student Succeeds Act*. www2.ed.gov/admins/lead/account/stateplan17/ksconsolidatedstateplanfinal.pdf

Kaplan, S. N. (2008). Projects: Yay or nay. *Gifted Child Today*, *31*(2), 47. https://doi.org/10.4219/gct-2008-759

Kaul, C. R., & Davis, B. K. (2018). How the state education agencies addressed gifted education in the Title II sections of their ESSA state plans. *Gifted Child Today*, *41*(3), 159–167. https://doi.org/10.1177/1076217518769700

Kettler, T. (2016). Why are economists evaluating the impact of gifted education? *Journal of Advanced Academics*, *27*(2), 81–89. https://doi.org/10.1177/1932202X16636172

Kim, K. H., VanTassel-Baska, J., Bracken, B., Feng, A., & Stambaugh, T. (2014). Assessing science reasoning and conceptual understanding in the primary grades using standardized and performance-based assessments. *Journal of Advanced Academics*, *25*(1), 47–66. https://doi.org/10.1177/1932202X14520946

Kitano, M. K. (2003). Gifted potential and poverty: A call for extraordinary action. *Journal for the Education of the Gifted*, *26*(4), 292–303. https://doi.org/10.4219/jeg-2003-305

Lee, S., & Olszewski-Kubilius, P. (2006). Comparison between talent search students qualifying via scores on standardized tests and via parent nomination. *Roeper Review*, *29*(3), 157–166. https://doi.org/10.1080/02783190609554355

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners: The contribution of dynamic assessment. *School Psychology International*, *22*(1), 74–96. https://doi.org/10.1177/01430343010221006

Lubinski, D. (2016). From Terman to today: A century of findings on intellectual precocity. *Review of Educational Research*, *86*(4), 900–944. https://doi.org/10.3102/0034654316675476.

Maxwell, M. (2007). Career counseling is personal counseling: A constructivist approach to nurturing the development of gifted female adolescents. *The Career Development Quarterly*, *55*(3), 206–224. https://doi.org/10.1002/j.2161-0045.2007.tb00078.x

McBee, M. T. (2016). What you don't look for, you won't find: A commentary on Card and Giuliano's examination of universal screening. *Journal of Advanced Academics*, *27*(2), 131–138. https://doi.org/10.1177/1932202X16634141

McCoach, D. B., Rambo, K. E., & Welsh, M. (2013). Assessing the growth of gifted students. *Gifted Child Quarterly*, *57*(1), 56–67. https://doi.org/10.1177/0016986212463873

Matthews, M. S., & Shaunessy, E. (2010). Putting standards into practice: Evaluating the utility of the NAGC Pre-K-Grade 12 Gifted Program Standards. *Gifted Child Quarterly*, *54*(3), 159–167. https://doi.org/10.1177/0016986209356708

Missett, T. C., & Foster, L. H. (2015). Searching for evidence-based practice: A survey of empirical studies on curricular interventions measuring and reporting fidelity of implementation published during 2004–2013. *Journal of Advanced Academics*, *26*(2), 96–111. https://doi.org/10.1177/1932202X15577206

Muratori, M. C., & Smith, C. K. (2015). Guiding the talent and career development of the gifted individual. *Journal of Counseling & Development*, *93*(2), 173–182. https://doi.org/10.1002/j.1556-6676.2015.00193.x

National Association for Gifted Children (2019a). *2019 Pre-k-grade 12 gifted programming standards*. www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20Standards%281%29.pdf

National Association for Gifted Children (2019b). *Programming standards glossary of terms*. www.nagc.org/programming-standards-glossary-terms

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. U. S. Department of Education. www.edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf

Newman, J. L. (2004). Talents and type IIIs: The effects of the talents unlimited model on creative productivity in gifted youngsters. *Roeper Review*, *27*(2), 84–90. https://doi.org/10.1080/02783190509554295

Olszewski-Kubilius, P., & Clarenbach, J. (2012). *Unlocking emergent talent: Supporting high achievement of low-income, high ability-students*. National Association for Gifted Children. www.nagc.org/sites/default/files/key%20reports/Unlocking%20Emergent%20Talent%20(final).pdf

Olszewski-Kubilius, P., & Kulieke, M. (2008). Using off-level testing and assessment for gifted and talented students. In J. VanTassel-Baska (Ed.). *Alternative assessments with gifted and talented students* (pp. 89–106). Prufrock Press.

Palmarini, J. (2011). The role of cornerstone tasks in the next generation of theatre standards: An interview with assessment expert Jay McTighe. Teaching Theatre. https://www.jaymctighe.com/wp-content/uploads/2011/04/Cornerstone-Tasks-in-Theater.pdf

Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, *60*(3), 159–171. https://doi.org/10.1177/0016986216643165

Reis, S. M., & Callahan, C. M. (Eds.). (2004). *Program evaluation in gifted education*. Corwin Press.

Renzulli, J. S., & Reis, S. M. (2003). The schoolwide enrichment model: Developing creative and productive giftedness. In N. Colangelo, & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 184–203). Allyn & Bacon.

Robinson, A., Cotabish, A., Wood, B. K., &O'Tuel, F. S. (2014). The effects of a statewide evaluation initiative in gifted education on practitioner knowledge, concerns, and program documentation. *Journal of Advanced Academics*, *25*(4), 349–383. https://doi.org/10.1177/1932202X14549356

Ryser, G. R. (2018a). Fairness in testing and nonbiased assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 59–72). Prufrock Press.

Ryser, G. R. (2018b). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 33–57). Prufrock Press.

Ryser, G. R., & Rambo-Hernandez, K. E. (2014). Using growth models to measure school performance. *Gifted Child Today*, *37*(1), 17–23. https://doi.org/10.1177/1076217513509617

Sanderson, E., & Greenberger, R. (2011). Evaluating online programs through a gifted lens. *Gifted Child Today*, *34*(3), 42–55.

Shin, M.-H. (2018). Effects of project-based learning on students' motivation and self-efficacy. *English Teaching, 73*(1), 95–114. https://doi.org/10.15858/engtea.73.3.201809.95

Siegle, D., Gubbins, E. J., O'Rourke, P., Langley, S. D., Mun, R. U., Luria, S. R., Little, C. A., McCoach, D. B., Knupp, T., Callahan, C. M., & Plucker, J. A. (2016). Barriers to underserved students' participation in gifted programs and possible solutions. *Journal for the Education of the Gifted*, *39*(2), 103–131. https://doi.org/10.1177/0162353216640930

Siegle, D., & McCoach, D. G. (2005). Extending learning through mentorships. In F. Karnes & S. Bean (Eds.), *Methods and materials for teaching gifted* (2nd ed., pp. 473–518). Prufrock Press.

Simonton, D. K. (2000). Cognitive, personal, developmental, and social aspects. *American Psychologist*, *55*(1), 151–158. https://doi.org/10.1037/0003-066X.55.1.151

South Dakota Department of Education. (2018, January 12). *South Dakota consolidated state plan: The elementary and secondary education act of 1965, as amended by the Every Student Succeeds Act.* www2.ed.gov/admins/lead/account/stateplan17/sdconsolidatedstateplanfinal2.pdf

Speirs Neumeister, K. L., Adams, C. M., Pierce, R. L., Cassady, J. C., & Dixon, F. A. (2007). Fourth-grade teachers' perceptions of giftedness: Implications for identifying and serving diverse gifted students. *Journal for the Education of the Gifted*, *30*(4), 479–499. https://doi.org/10.4219/jeg-2007-503

Speirs Neumeister, K., & Burney, V. H. (2019). *Gifted program evaluation: A handbook for administrators and coordinators* (2nd ed.). Prufrock Press.

Sriraman, B. (2004). Gifted ninth graders' notions of proof: Investigating parallels in approaches of mathematically gifted students and professional mathematicians. *Journal for the Education of the Gifted*, *27*(4), 267–292. https://doi.org/10.4219/jeg-2004-317

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K-12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102/0034654316675417

Subotnik, R. F., Almarode, J., & Lee, G. M. (2016). STEM schools as incubators of talent development. *Gifted Child Today*, *39*(4), 236–241. https://doi.org/10.1177/1076217516661592

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Sulak, T. N. (2014). Using CBM to identify advanced learners in the general education classroom. *Gifted Child Today*, *37*(1), 25–31. https://doi.org/10.1177/1076217513509620

Swanson, J. D. (2006). Breaking through assumptions about low-income, minority gifted students. *Gifted Child Quarterly*, *50*(1), 11–25. https://doi.org/10.1177/001698620605000103

Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, *71*(2), 321–363. https://doi.org/10.3102/00346543071002321

Thompson, D. D., & McDonald, D. M. (2007). Examining the influence of teacher-constructed and student-constructed assignments on the achievement patterns of gifted and advanced sixth-grade students. *Journal for the Education of the Gifted*, *31*(2), 198–226. https://doi.org/10.4219/jeg-2007-676

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29*(1), 60–89. https://doi.org/10.1177/016235320502900104

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of the literature. *Journal for the Education of the Gifted*, *27*(2–3), 119–145. https://doi.org/10.1177/016235320302700203

U.S. Department of Education (2015). *The Every Student Succeeds Act*. www.ed.gov/essa

U.S. Department of Education. (2017). *ESSA state plan submission*. www2.ed.gov/admins/lead/account/stateplan17/statesubmission.html

van Deur, P. (2011). Views of gifted elementary students about self-directed learning. *Gifted and Talented International*, *26*(1/2), 111–120.

VanTassel-Baska, J., & Baska, A. (Eds.). (2020). *Curriculum planning and instructional design*. Prufrock Press.

VanTassel-Baska, J., & Hubbard, G. F. (2019). A review of the national gifted standards implementation in eight districts: An uneven picture of practice. *Gifted Child Today*, *42*(4), 215–228. https://doi.org/10.1177/1076217519862336

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 gifted programs: A clarion call for enhanced gifted program development. *Gifted Child Quarterly*, *50*(3), 199–210. https://doi.org/10.1177/001698620605000302

VanTassel-Baska, J. (Ed.). (2009). *Patterns and profiles of promising learners from poverty*. Prufrock Press.

VanTassel-Baska, J. (2014). Performance-based assessment: The road to authentic learning for the gifted. *Gifted Child Today*, *37*(1), 41–47. https://doi.org/10.1177/1076217513509618

VanTassel-Baska, J., Avery, L.D., Little, C., & Hughes, C. (2000). An evaluation of the implementation of curriculum innovation: The impact of the William and Mary units on schools. *Journal for the Education of the Gifted*, *23*(2), 244–272. https://doi.org/10.1177/016235320002300201

VanTassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly*, *51*(3), 218–231. https://doi.org/10.1177/0016986207302717

VanTassel-Baska, J., Johnson, D., & Avery, L. D. (2002). Using performance tasks in the identification of economically disadvantaged and minority gifted learners: Findings from Project STAR. *Gifted Child Quarterly*, *46*(2), 110–123. https://doi.org/10.1177/001698620204600204

Walker, L. D., & Vander Ploeg, M. K. (2015). Surveying graduates of a self-contained high school gifted program: A tool for evaluation, development, and strategic planning. *Gifted Child Today*, *38*(3), 160–176. https://doi.org/10.1177/1076217515583741

Warne, R. T. (2014). Using above-level testing to track growth in academic achievement in gifted students. *Gifted Child Quarterly*, *58*(1), 3–23. https://doi.org/10.1177/0016986213513793

Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*, *66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009

Yoon, S.Y., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*, *53*(2), 121–136. https://doi.org/10.1177/0016986208330564

PART

I

Identification

Introduction to Part I on Assessments for Identification

Susan K. Johnsen and Joyce VanTassel-Baska

Introduction to Identification

Identification is an ongoing process of gathering information by using appropriate tests, instruments, and techniques. In gathering information, educators need to develop environments and learning experiences that encourage all students to express their gifts and talents, use multiple sources of information, select qualitative and quantitative instruments in different domains, and consider multiple indicators of potential and achievement that relate to student characteristics and programming options (NAGC, 2019).

Purposes of Identification

The purpose of identification is to find gifted and advanced learners who require modification to their programming to optimize their learning. Without identification as a precursor to these modifications, gifted and advanced learners may experience adverse effects, which is particularly true for students of color and those from poverty and other underserved populations (Olszewski-Kubilius & Corwith, 2018; Plucker et al., 2013; Xiang et al., 2011). The importance of identifying gifted and advanced learners students for specialized services cannot be underestimated in terms of student future development. Educators need to examine assessment tools and procedures that minimize bias so that they have more opportunities to recognize students from all backgrounds.

Models of Identification

According to the National Association for Gifted Children Pre-K–Grade 12 Gifted Programming Standards (NAGC, 2019), identification models need to be comprehensive, cohesive, and ongoing (see Standard 2.2.1). *Comprehensive* means that students are identified across all grade levels—preschool to grade 12; *cohesive* means that students

DOI: 10.4324/9781003285991-3

receive services and programs that are aligned with their interests, strengths, and needs; and *ongoing* means that students might be identified or discovered as they display characteristics over the time they are in school.

To meet this standard, educators must select and use assessments that are aligned with the students' characteristics and need for services. The process can be complicated by the diversity of the gifted population, which include students with a variety of gifts in a variety of domains, who are from low socioeconomic and different cultural backgrounds, and who display uneven development. Assessments therefore need to be carefully selected to ensure that students receive appropriate services and programs. These assessments are frequently embedded within one or more phases (e.g., nomination/referral, screening, selection).

In more inclusive models, all of the assessments are administered to all of the students at a particular grade level whereas in other models, assessments may be administered only after a student has been referred. McBee et al. (2016) argue that a more inclusive model is more likely to find students from underrepresented groups because of educators who may not be aware of a broad range of characteristics or may have implicit biases. Mun et al. (2020) agree that cultural bias may be present in excluding some students, particularly those who are English learners, but they suggest a more multi-phase process that uses a strengths-based approach, beginning with observations of students' thinking strategies followed by the use of multiple measures and alternative assessments. In considering twice exceptional students (i.e., gifted students with disabilities), Crepeau-Hobson and Bianco (2010) have proposed integrating identification of gifted students within a response-to-intervention model. Within this model, Tier 1 includes universal screening; Tier 2 provides interventions with differentiated instruction, acceleration, and enrichment; and Tier 3 uses comprehensive psychoeducational assessments, culminating in an Advanced Learning Plan.

Identification Questions

Regardless of the model used, educators need to consider these specific questions in developing an identification process.

What Are the Characteristics of the Students We Are Identifying?

Characteristics of gifted and advanced students have been identified in the research literature for general intellectual ability, specific academic fields, creativity, visual and performing arts, leadership, and affective domains (Piirto, 2007; Renzulli et al., 2010; Rimm et al., 2018; Ryser et al., 2021). Many general characteristics like early reading, working with puzzles, and power of concentration may be supplemented by more specific characteristics like solving advanced math problems, building a complex system, or creating a story worthy of publication. Most checklists for teacher and parent identification of giftedness focus on specific characteristics that demonstrate the need for advanced opportunities. Affective characteristics also figure prominently in decisions about the giftedness of children. Their humor, their empathic concern for others in the universe including animals, and their perfectionism all present as characteristics of note. In addition, characteristics of underrepresented groups such as students with disabilities and those from lower socioeconomic and culturally diverse backgrounds have also been identified (Baum et al., 2017; Esquierdo & Arreguín-Anderson, 2012; McCoach & Siegle, 2003; VanTassel-Baska & Stambaugh, 2007). All educators need to be familiar with these characteristics as they begin the identification planning process.

What Assessment Tools Align to the Students' Characteristics?

Knowing characteristics of gifted and advanced students is an important first step in identification. Next, the focus should turn to the assessment tools that will help educators find students who possess such characteristics. For example, if students are English learners, the assessments would need to be in the student's native language or examined for language requirements. In this case, nonverbal assessments might need to be considered. If a student is performing at a higher grade level, then tools with a sufficient ceiling or above-level should be selected. Students with disabilities will also need to be administered assessments accommodating both their needs and their strengths. Moreover, students from lower socioeconomic backgrounds may need assessments where test items do not require prior access to formal educational opportunities. In all cases, the assessment's manual should describe how the test is able to identify students from a variety of diverse backgrounds.

What Are the Plans for Services and Programs Based on Student Characteristics?

Given the variation in strengths, needs, and interests among gifted and advanced students, educators need to plan a comprehensive, cohesive, and coordinated array of services. These services may include talent development activities prior to identification, acceleration, grouping, individualized learning options, and enrichment options within and outside of the school setting (NAGC, 2019; Siegle et al., 2016; Steenbergen et al., 2016). Learning progressions need to be developed beginning in kindergarten through grade 12 with transition support for post-secondary options. Moreover, all educators who provide gifted and advanced, general, special, and related professional services need to be involved to ensure a coordinated set of programming supports (Bianco & Harris, 2014; Coleman & Hughes, 2009). Characteristics of students should be considered in developing educational provisions prior to any form of identification, but the general goals and outcomes of planned gifted and advanced programming can be identified. Both student characteristics and programming options will influence the selection of assessments.

What Sources and Types of Assessments Will We Select?

Both qualitative and quantitative approaches to assessment should be considered in identifying students (NAGC, 2019; Ryser, 2018b; Worrell, 2009). Qualitative assessments use words whereas quantitative measures use numbers to describe the student's strengths, needs, and interests. While quantitative assessments (e.g., aptitude tests, intelligence tests, achievement tests, rating scales) are norm-referenced and standardized, qualitative assessments (e.g., performance- and product-based, observations, dynamic learning activities) are not but provide greater flexibility and richer descriptions of the student's actual performance. When different sources are added to the assessments (e.g., teacher, parent/guardian, peers, the student), a more comprehensive view of the student's strengths, needs, and interests in different contexts can be identified and used to make decisions.

How Do We Ensure Technically Adequate Assessments?

In assessing technical qualities, educators need to review the validity and reliability of both qualitative and quantitative assessments and examine the technical manual of the standardized, norm-referenced tests (Robins & Jolly, 2018; Ryser, 2018a). In the area of reliability, evaluators should consider consistency (i.e., Is the instrument assessing the same construct? Does the instrument have error?), stability (i.e., Is the performance stable over time? Do samples of work vary across teachers and time?), and inter-rater (i.e., Would two scorers evaluate the evidence similarly?). In the area of validity, evaluators should consider the content (i.e., Does the product/performance or test items relate to the content being measured such as math, literacy, the arts?), criterion-related (i.e., Does the assessment relate to other assessments that have similar content or future performance?), and the construct underlying the assessment (i.e., Does the test have studies that support underlying hypotheses, such as factors relating to the content or gifted students performing better than comparison groups?).

Equally important are issues related to the norming population and bias of tests. In the case of norming, the test should show whether or not the norms resemble national census data. If the school or school district's population is different from the national norms, the evaluator can then decide if local norms should be developed and used (Gray et al., 2009). In the case of bias, along with the representativeness of the norms, the assessment should show that groups perform similarly on different items on the test (e.g., items are familiar to some students but not others; different subgroups have the same probability of answering an item correctly) and that language demands do not interfere with the subgroup's performance (Ryser, 2018a).

Any standardized, norm-referenced assessment instrument needs to have a technical manual that reports these data. Better yet, has the instrument been used in research studies where these data are routinely reported? A district coordinator must compile the technical data for review by the committee responsible for selecting the instrumentation that includes relevant manuals and studies that have used the instruments under consideration.

What Process Will Be Used in Implementing the Assessments?

To implement the identification process with fidelity, policies and procedures need to be developed that include referral, informed consent, the assessment process, review of assessment information, student retention, student reassessment, student exiting, and appeals (see NAGC evidence-based practice 2.2.1). The policies may also include approaches that minimize bias such as front-loading, talent development activities, local norms, nonverbal assessments, and building relationships (see NAGC evidence-based practice 2.3.1).

Schools will need to consider how many phases will be used. Will the school administer all of the assessments to all of the students and then review for placement? Will there be a referral process before any tests are administered? How many tools should be employed? Several states require three or more assessments. If multiple assessments are used, is each instrument equally valuable? Some states weight instruments in terms of their value to the process; others provide cutoff scores on each instrument for inclusion of students. What will be the mix of quantitative measures like tests and more qualitative assessments like products or performances? A clear answer to these questions is needed to construct an identification system that will yield those students who need services beyond the general education program.

How Will the Assessments Be Interpreted?

After all of the information is collected, a committee of individuals who understand the characteristics of gifted and advanced students and their need for different types of services need to review *all* of the assessment information (Johnsen, 2018b; Moon, 2013). If the data are quantitative, the committee needs to decide if local or national norms will be used. This decision is most likely based on the population of the entire school district or individual campuses. Scores also need to be comparable (e.g., standard/index scores to standard/index scores), and test error should be included. For example, suppose that a student scores 120 (91st percentile) on an intelligence test, and the standard error is 3 points. The interpreter of this score could say that 68% of the time, the student would score in the range of 117–123; 95% of the time in the range of 114–126; and 99% of the time in the range of 111–129. Error is particularly important when schools consider strict cut-off scores. Many districts use "handmade" tools for checklists of characteristics which should be replaced with more standardized ones that have technical data.

Qualitative data such as anecdotal observations, responses to dynamic assessments, and products and performances are particularly helpful in matching services to students' interests, strengths, and needs. A system for collecting student products and assessing them by a neutral committee is critical. The student's best performance may show potential and be weighed to identify. All of the information should be reviewed and considered. Special attention should be given to those students who are underrepresented to ensure equity. The most important purpose in making decisions about placements is to ensure that all students receive the programming and services they need to realize their potential.

What Professional Learning Needs to Be Provided? To Whom?

Once all of the policies and procedures are developed and the assessments are selected, all educators need to receive professional learning, specifically about each tool selected, why, and what information it provides, and the identification processes. The professional learning will involve administrators, general educators, gifted educators, special educators, counselors, and other instructional support staff so they might collaborate in not only implementing the identification process but also in providing appropriate programming and services. With professional learning, educators are able to observe high performance (Pendarvis & Wood, 2009) and minimize bias (Olszewski-Kubilius, & Corwith, 2018).

Professional learning about the identification process should follow the outline of topics described in this section. Teachers and parents, in separate sessions, should review the characteristics of gifted and advanced children, practice the use of the checklist to be used and discuss issues that arise over interpretations of items. Teachers should be provided technical adequacy data and other info about all instruments, including results of studies where they have been used. Teachers should also review why they are using a given test, what it purports to say about a student's advanced ability and what the implications are for curriculum matching.

How Should Parents/Guardians Be Involved with the Identification Process?

Similar to the participation of educators in professional learning, parents/guardians need to be oriented to the importance of the identification process and appropriate services for their children. Parent checklists may be used to codify their understanding of their child's characteristics and needs. Moreover, educators need to develop relationships and build home/school connections, which appears to be critical in identifying more students from culturally, linguistically, and ethnically diverse gifted and advanced students (Briggs et al., 2008). Once parents/guardians understand the characteristics of gifted and advanced students, they can be involved in sharing observations and information about their children in a different context. Their involvement also improves their understanding of assessments and related programming needs, higher expectations, support, and the ultimate success of their students (Kaul et al., 2015; Wilder, 2014).

How Will the Effectiveness of the Identification Process Be Evaluated?

One of the best ways to ensure that the identification system is effective is to conduct predictive validity studies on the instruments used. Did they identify the right students? How successful were students who were identified for the program? Could more students be identified for the program, given the overall numbers identified and the level of performance of the pool? How well did the tools work together in providing different information that is useful for programming and services? Of great importance is also how teachers actually used the results for planning instruction. Did it help them do more differentiation within the group? Did it help to plan individual opportunities or cluster projects?

Johnsen (2018a) describes how an evaluation can ensure that the identification procedure is aligned to state and national standards, implemented with fidelity and meeting the desired outcomes (p. 151). With such an evaluation, educators are able to determine which components are effective for identifying those students who need programming and services within and beyond the general education program. Recommendations can also be made about components that need to be changed so that all eligible students are appropriately served.

Overview of Chapters

The chapters in this section provide readers with additional information related to many of these questions. In the first chapter, Lindsay Lee and Scott Peters describe the design of universal screening and universal consideration systems and suggest that these models improve equity. The second and third chapters address quantitative and qualitative assessments. Using the structure of human abilities, Lakin, Simonds, and Caviness-French identify the key measurement concepts when selecting and interpreting different types of quantitative assessments. Focusing on qualitative assessments, Little, O'Brien, and Kearney examine curriculum-based and performance tasks and their use in identification systems. The next chapter investigates specific types of instruments commonly used in the initial phases of identification—teacher rating scales. Ryser describes their important characteristics and how they complement information from other measures. The final chapter in this section examines the critical area of identifying gifted students from underrepresented populations. Through their review of the literature and interviews with coordinators from diverse districts, VanTassel-Baska and Johnsen provide information about effective instruments and processes and principles for reducing disparities.

Conclusion

This introduction to the identification section of the book has provided a blueprint for the central questions that the following chapters address. The importance of understanding the characteristics of the populations of advanced learners, coupled with the processes that are most likely to yield results in finding them, form the crux of the chapters in this section. Promising instruments, approaches to finding these students, and the use of other tools are reviewed along with the ways in which they are implemented in schools to find both typical gifted and advanced learners and more underrepresented groups. Both traditional and non-traditional approaches are shared that have research on effectiveness, including tests, teacher and parent inventories, student products and performances. A model for practitioner use concludes the section as best practice guidelines become reality in school-based plans.

References

Baum, S. M., Schader, R. M., & Owen, S. V. (2017). *To be gifted and learning disabled: Strength-based strategies for helping twice-exceptional students with LD, ADHD, ASD, and more* (3rd ed.). Prufrock Press.

Bianco, M., & Harris, B. (2014). Strength-based RtI: Developing gifted potential in Spanish-speaking English language learners. *Gifted Child Today*, *37*(3), 169–176. https://doi.org/10.1177/1076217514530115

Briggs, C. J., Reis, S. M., & Sullivan, E. E. (2008). A national view of promising programs and practices for culturally, linguistically, and ethnically diverse gifted and talented students. *Gifted Child Quarterly*, *52*(2), 131–145. https://doi.org/10.1177/0016986208316037

Coleman, M. R., & Hughes, C. E. (2009). Meeting the needs of gifted students within an RtI framework, *Gifted Child Today*, *32*(1), 14–17. https://doi.org/10.1177/107621750903200306

Crepeau-Hobson, F., & Bianco, M. (2010). Identification of gifted students with learning disabilities in a response-to-intervention era. *Psychology in the Schools*, *48*(2), 102–109. https://doi.org/10.1002/pits.20528

Esquierdo, J. J., & Arreguín-Anderson, M. (2012). The "invisible" gifted and talented bilingual students: A current report on enrollment in GT programs. *Journal for the Education of the Gifted*, *35*(1), 35–47. https://doi.org/10.1177/0162353211432041

Gray, R., McCallum, R. S., & Bain, S. K. (2009). Language-reduced screening for giftedness. *Journal for the Education of the Gifted*, *33*(1), 38–64. https://doi.org/10.1177/016235320903300103

Johnsen, S. K. (2018a). Evaluating the effectiveness of identification procedures. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 151–160). Prufrock Press.

Johnsen, S. K. (2018b). Making decisions about placement. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 117–150). Prufrock Press.

Kaul, C. R., Johnsen, S. K., Witte, M. M., & Saxon, T. F. (2015). Critical components of a summer enrichment program for urban low-income gifted students. *Gifted Child Today*, *38*(1), 32–40. https://doi.org/10.1177/1076217514556533

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

McCoach, D. B., & Siegle, D. (2003). Factors that differentiate underachieving gifted students from high-achieving gifted students. *Gifted Child Quarterly*, *47*(2), 144–154. https://doi.org/10.1177/001698620304700205

Moon, T. (2013). Uses and misuses of matrices in identifying gifted students: Considerations for better practice. In C. M. Callahan, & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (pp. 128–134). Routledge

Mun, R. U., Hemmler, V., Langley, S. D., Ware, S., Gubbins, E. J., Callahan, C. M., McCoach, D. B., & Siegle, D. (2020). Identifying and serving English learners in gifted education: Looking back and moving forward. *Journal for the Education of the Gifted*, *43*(4), 297–335. https://doi.org/10.1177/0162353220955230.

National Association for Gifted Children (NAGC). (2019). *NAGC 2019 pre-K-grade 12 gifted programming standards*. www.nagc.org/resources-publications/resources/national-standards-gifted-and-talented-education/pre-k-grade-12

Olszewski-Kubilius, P. & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review, *Gifted Child Quarterly*, *62*(1), 37–55. https://doi.org/10.1177/0016986217738015

Pendarvis, E., & Wood, E.W. (2009). Eligibility of historically underrepresented students referred for gifted education in a rural school district: A case study. *Journal for the Education of the Gifted*, *32*(4), 495–514. https://doi.org/10.1177/016235320903200403

Piirto, J. (2007). *Talented children and adults: Their development and education* (3rd ed.). Prufrock Press.

Plucker, J. A., Hardesty, J., & Burroughs, N. (2013). *Talent on the sidelines. Excellence gaps and America's persistent talent underclass*. www.nagc.org/sites/default/files/key%20reports/Talent%20on%20the%20Sidelines%20(2013).pdf

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., Gavin, M.K., Reis, S. M., Siegle, D., & Stysma Reed, R. E. (2010). *Scales for rating the behavioral characteristics of superior students* (3rd ed.). Prufrock Press.

Rimm, S. B., Siegle, D., & Davis, G. A. (2018). *Education of the gifted and talented* (7th ed.). Pearson.

Robins, J., H., & Jolly, J. L. (2018). Technical information regarding assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 73–116). Prufrock Press.

Ryser, G. R. (2018a). Fairness in testing and nonbiased assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 59–72). Prufrock Press.

Ryser, G. R. (2018b). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 33–57). Prufrock Press.

Ryser, G. R., McConnell, K., Sanguras, L.Y., & Kettler, T. (2021). *Scales for identifying gifted students* (2nd ed.). Prufrock Press.

Siegle, D., Gubbins, E. J., O'Rourke, P., Langley, S. D., Mun, R. U., Luria, S. R., Little, C. A., McCoach, D. B., Knupp, T., Callahan, C. M., & Plucker, J. A. (2016). Barriers to underserved students' participation in gifted programs and possible solutions. *Journal for the Education of the Gifted*, *39*(2), 103–131. https://doi.org/10.1177/0162353216640930

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K-12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102/0034654316675417

VanTassel-Baska, J., & Stambaugh, T. (2007). *Overlooked gems: A national perspective on low-income promising learners*. National Association for Gifted Children. www.nagc.org/sites/default/files/key%20reports/Overlooked%20Gems%20%28final%29.pdf

Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*, *66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009

Worrell, F. C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly*, *53*(4), 242–244. https://doi.org/10.1177/0016986209346828

Xiang, Y., Dahlin, M., Cronin, J., Theaker, R., & Durant, S. (2011). *Do high flyers maintain their altitude? Performance trends of top students*. Fordham Institute. https://fordhaminstitute.org/national/research/do-high-flyers-maintain-their-altitude-performance-trends-top-students.

CHAPTER

2

Universal Screening

A Process to Promote Equity

Lindsay Ellis Lee and Scott J. Peters

Introduction

Who gets into gifted programs is of continual interest to gifted and talented researchers and administrators. Partly, this is due to the continued underrepresentation of students who are Black, Latinx, from low-income families, still learning English, or twice-exceptional (Grissom et al., 2019; Peters et al., 2019). Universal screening has become a popular method to miss fewer students and improve equity. For instance, Card and Giuliano (2016) found this was the case when a large Florida district moved from selective screening to universal screening. After implementing universal screening, program participation increased to 5.2% (from 3.5%) and improved equitable access across schools. Not only did the identification improve for all students, but it increased the most for Black and Latinx students. However popular this method has become as a process to promote equity, there are still misunderstandings on how to effectively implement universal screening, nor is there enough attention to systems of universal consideration where all students are put through the full identification process. Thus, this chapter aims to define universal screening in the context of gifted education and how procedures are designed with a particular focus on the construction of identification systems. Further, we will discuss how universal screening can be used to maximize sensitivity, minimize cost, and improve equitable access to gifted services.

What is Universal Screening?

Screening refers to triaging a group to see who should move on to the second phase of more in-depth evaluation. Screening systems can be *selective*, where only some portion of a larger population is screened, or they can be *universal*, where everyone in a population undergoes the process. For instance, universal screeners are a typical first step in data collection for multi-tiered systems, such as Response-to-intervention (RTI), that aim to address early prevention and intervention for students *at-risk* in academic domains, or even address social, emotional, and behavioral needs of students (Romer et al., 2020). Typically, universal screeners have been used to evaluate whether students are "on track" in their competency and skills, monitor the progress of interventions, and

DOI: 10.4324/9781003285991-4

measure desired outcomes for students. Although screening tools in schools are commonly associated with finding where students are struggling, screeners can also be used to find student strengths and talents, as seen in screeners for gifted programs.

Universal Screening in Gifted Education

In the context of gifted education, *universal screening* consists of screening students with a specific aptitude and/or achievement measure to determine who should enter the formal identification process (i.e., the second two-phase identification system). This screening most often occurs within a specified grade level or at multiple time points in elementary and/or secondary. Although it typically occurs in a single grade level, there are no set "rules" that dictate its occurrence at a single time point, besides the time, money, and training it takes to implement. The assessments used are commonly quantitative (e.g., achievement, ability tests), however, some districts have implemented qualitative tools (e.g., rating scales, observations, portfolios). Nothing about universal screening dictates the particular instrument to be used; it is simply the practice of giving the selected instrument to all students as a way to determine who should move on to formal evaluation or the identification phase.

Alternatively, *universal consideration* is a single-phase identification system where all students, often at a single grade level, are put through the full process for gifted service eligibility. For example, imagine a district that makes placement decisions based on three data points: the Cognitive Abilities Test (CogAT), the Measure of Academic Progress (MAP), and teacher ratings of gifted characteristics. If these data points are collected to make identification decisions from *all* students in a given grade, then this is an example of universal consideration. This can also be thought of as a single-phase identification system because there are not multiple phases for who gets considered. Instead, the entire population of students is considered.

Multi-Phase Identification Systems

Two-Phase Identification Systems: Universal Screening

As described above, universal screening is the first phase of a two-phase system of gifted identification whereby all students of a given population (e.g., a single grade level) are considered to determine who should be formally evaluated in Phase Two. Students who pass through this screening process are then put through the formal service eligibility determination process. Two-phase systems are a much more common gifted identification process than universal consideration for reasons we will talk about below. *Two-phase identification systems* can take one of two forms. In the first, all students are put through a universal screening process, which then determines who moves on to the second phase. For example, perhaps all second-grade students take the CogAT screener to determine who should take the full CogAT—the instrument that determines service eligibility. In the second, students are not universally screened, but instead must receive a referral of some kind (e.g., nominations from teachers, parents, community members, self) to proceed to the second phase. We refer to this as *selective screening*. The key difference is that in the first option, all students are put through some kind of screening, whereas in the second option, not only are some students never considered, but many students will not even be screened since they were never referred or nominated. See Figure 2.1 for a

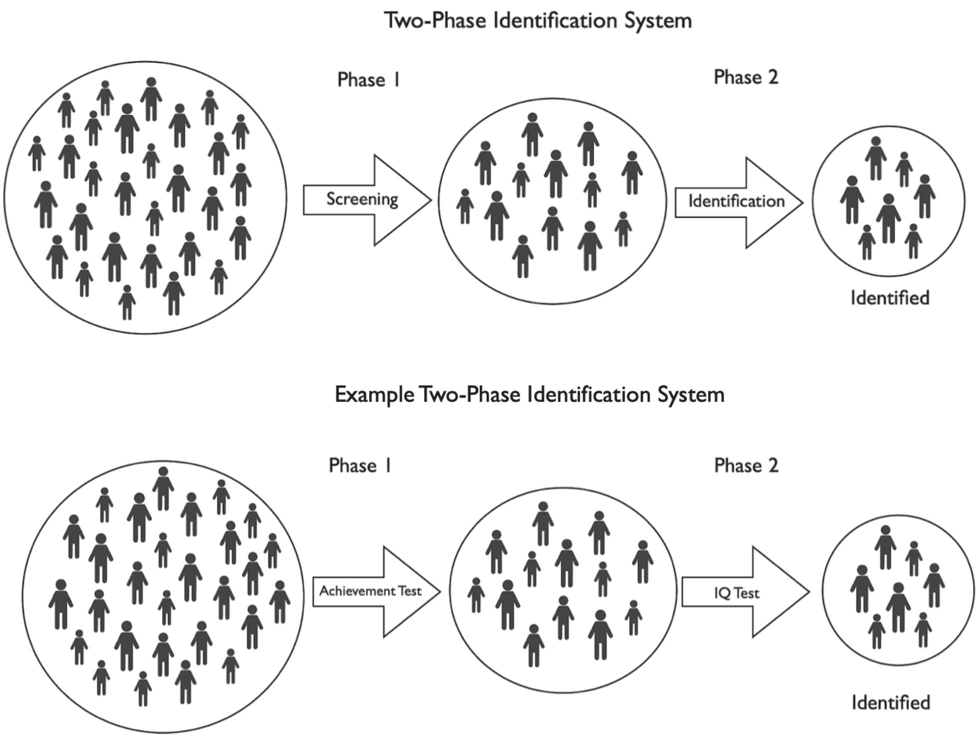

FIGURE 2.1 Two-phase identification system

conceptualization of a two-phase system that uses universal screening and an example of a two-phase identification system.

This brings us to the concept of sensitivity in the identification process. Sensitivity is the percentage of students the identification system correctly identifies. McBee et al. (2016) describe variables that come into play when estimating the sensitivity of any two-phase process:

1. The phase-one cut scores determine who moves on to phase-two
2. The reliability of the phase-one screening assessment(s)
3. The reliability of the phase-two identification assessment(s)
4. The phase-two cut score determines service eligibility
5. The nomination validity (correlation between the two phases).

Two-phase systems are also common in medical testing. Early on in the COVID-19 Pandemic, students in schools or colleges were not being universally tested because there were not enough tests available—it was an issue of limited resources. Instead, students were often asked a set of screening questions such as whether they had been in contact with anyone who was ill or if they had lost their sense of smell or were having any other symptoms. If the answer was yes, then they were tested. The test itself made the diagnosis while the screening questions determined who should be tested.

Although two-phase systems are quite common, it is important to emphasize that two-phase systems, when compared to single-phase systems, *can never make sensitivity* *better*. They can only miss *more* students. However, when implemented correctly, they can generate similar sensitivity at far less cost. As seen in Figure 2.1, consider a hypothetical identification system where all students take an achievement test (Test reliability = .95). Students who score at the 90th percentile are formally considered for gifted services in Phase Two. This second phase involves an IQ test (Test reliability = .95). Students who score at the 95th percentile are identified as gifted. The two phases are correlated at .70. In other words, the high reliability of the instruments used for screening and the strength of the relationship between these instruments directly relates to how sensitive a system is for identifying students for appropriate placement. Using methods, we will explain in more detail below (also see McBee et al., 2016), this system has a sensitivity of .51, or just over half of the students being correctly identified.

A Single-Phase Identification System: Universal Consideration

Contrast the two-phase system described above with a single-phase system whereby *all* students in a single grade level take the same IQ test (Test reliability = .95) and must score at the 95th percentile to be identified as gifted. That system, with the same "gifted" test and cut score, results in a sensitivity of .82, meaning 82% of students being correctly identified. Why the stark difference between the two? Due to the two-phase universal screening system, many students who would have scored well enough on the IQ test to be identified were never given the test because they did not score high enough on the screening phase. Essentially, they did not score high enough to get over the hurdle between Phases One and Two. Universal consideration erases this factor. The nomination cut score, reliability of the nomination assessment(s), and the nomination validity are non-factors because they do not exist! Instead, all students go through the entire identification process that determines service eligibility. The result is an identification system that is limited solely by the reliability of the assessment(s) and the cut score used to determine eligibility.

Unfortunately, the benefits of universal consideration come at extreme costs. A school district of 5000 students in grades K-12 might have around 400 in second grade—the grade during which the district conducts universal screening. If each IQ test takes two hours to administer, then administering these tests alone will take 800 hours, equivalent to a half-time employee. This does not include the variable costs beyond the psychologist's time. But regardless of the actual dollar amount, universal consideration is extremely inefficient. This is because 100% of students are being tested even though only a small percentage (by definition) will qualify. Cost and inefficiency is no mere academic issue. Card and Giuliano (2016) described a district that was forced to discontinue a similar identification because it was too costly. Further, every minute and dollar spent on assessment and identification is a minute or dollar not spent on direct service delivery. For this reason, identification systems need to be as efficient and effective (sensitive) as possible.

What Makes a "Good" Universal Screener?

No instrument is an inherently "good" or "bad" universal screener. However, there are a few simple rules when it comes to selecting a universal screener as part of a two-phase identification process. But first, it is worth re-emphasizing that universal screening presumes it will be given to all students of a given population (hence universal)—such as all second- or fifth-grade students. Absent all students taking the phase-one assessment, the result is a *selective* screening system where a human decision determines which students get referred for screening. This makes for a pseudo-three-phase system where Phase Three determines service eligibility, Phase Two is the referral for testing, and Phase One is some kind of individual, subjective decision on whether or not to refer in the first place. Hopefully, it is immediately clear that three-phase systems are a bad idea. Every additional phase is an additional hurdle for students to jump over and increases more opportunities for students to be missed. This is why the first essential criteria for a good or effective phase-one assessment is that it is a universally administered screener and not one that is restrictive in practice by being given only to some students.

Alignment of the Universal Screener to Program

School leadership should choose a screener that aligns with the intended program or intervention to be provided. This requires moving beyond the conceptualization of "giftedness" itself and finding the "truly gifted" (Peters & Borland, 2020) to considering the specific service. For instance, the use of a non-verbal screener for a program focused on enrichment in English Language Arts would not be in alignment at the program level, and especially would not be appropriate if high performance on a non-verbal ability test (Phase One) was required before administering a verbal ability test and reading achievement test (Phase Two) for identification. It is also vital to note in this scenario, the mere usage of a non-verbal screener in the process does not guarantee equitable access and could contribute to inequitable outcomes if improperly used. Thus, universal screeners should align with the specific program and administrators need to consider the details for how the screener is used (Lee et al., 2020).

Beyond alignment, other criteria include that the screener(s) is (a) time-efficient (e.g., quick to administer and for students to take), (b) cost-efficient (i.e., as inexpensive as possible), (c) strongly related to performance on the phase-two assessments (i.e., nomination validity), and (d) highly reliable. For example, district-created checklists as universal screeners could be viewed by administrators as "cost-efficient" but run the risk of not being reliable or accurate in determining who could be successful in Phase Two importantly, the program itself. We address each of these in turn, followed by an example of an effective two-phase system.

Time and Money

There is no point in implementing a two-phase identification system if the universal screener is as expensive and time-consuming as the phase-two assessment(s). After all, the only reason to use a two-phase system is to save time and money. If the phase-two assessments cost $25 per student and two hours to administer, but the universal screener costs $27 per student and five hours to administer, then there is little benefit. The district would be better off implementing a universal consideration system. This is why any universal screener, whether made up of one assessment or several data points, needs quick administration and inexpensive costs as possible. So, the screener should be easier, faster, and cheaper than the actual phase-two assessments that determine service eligibility.

In following these guidelines, one of the best universal screeners from a cost and time perspective is also used as part of the identification process at Phase Two. For example, if a district uses a group-administered cognitive ability test, a standardized achievement test, and a teacher rating scale (all entered into a matrix) to make identification decisions, then an ideal universal screener is one of those three data points. The most likely candidate is the academic achievement test that schools often already administer for other instructional purposes. As such, the cost and additional time required for this as a screener are essentially zero because the test is already being administered. Because ⅓ of the data in Phase Two is the same as the data in Phase One, the nomination validity (i.e., the correlation between the phases) is going to be higher than if a universal screener were used that was not also part of Phase Two.

In summary, a good universal screener must be inexpensive (at least less expensive than phase-two assessment[s]) and have quick administration in terms of teacher and student time. One of the best ways to satisfy these criteria is to use an assessment that is already being given to all students for some other instructional process (e.g., a universally administered achievement test) since using it as a screener would require no additional time or cost. That said, it is still possible to implement a two-phase system even if you are not fortunate enough to have something that all students have already taken.

Nomination Validity

Conceptually, nomination validity is how related each phase is to the other in a two-phase identification system. Ideally, we want all of the students who would do well on Phase Two to also do well on Phase One (and vice versa). Technically, nomination validity is the correlation between Phases One and Two. It answers the question of "how valid of a screener is Assessment X for a certain phase-two identification criteria?." Nomination validity is harder to estimate than something like time or cost. The primary place to look for information about the correlation between two assessments is in test manuals. For example, the technical manual for the Scales for Identifying Gifted Students (SIGS: Ryser & McConnell, 2004) reports a correlation between the SIGS general intellectual subscale and the composite of the CogAT Form 6 of .48. Similarly, the SIGS creativity subscale is correlated with the Torrance Test of Creative Thinking—Figural (TTCT: Torrance, 1990) at .62. In both cases, the SIGS would not be a strong universal screener for the CogAT or the TTCT because the nomination validity coefficients are low. The two phases may not be measuring the same constructs. It is possible to compensate for low nomination validity by setting a lower phase-one cut score (e.g., 80–90th percentile), but this means more students being tested at Phase Two, which in turn adds time and cost. This is why an effective two-phase identification system should select a Phase One universal screener that is as strongly correlated with the second phase as possible.

High Reliability

A final consideration for an effective universal screener is for it to yield highly reliable data. This technical information can be found in instrument test manuals for published assessment instruments. For example, Level B of the Naglieri Nonverbal Ability Test (Third Edition: Naglieri, 2018) showed internal consistency reliability of .88. The Fountas and Pinnell Benchmark Assessment System (F&P: Heinemann, n.d.) demonstrated test-retest reliability of about .94. Strong reliability is important at both phases and will result in fewer students being missed and a stronger, more sensitive identification system being created overall.

Even if it is not possible to look up or calculate the reliability of an assessment, there are some guiding general rules for reliable assessments. Most importantly, informal, teacher-made, or subjectively-scored assessments (e.g., portfolios, essays) will be less reliable than something like multiple-choice standardized tests. Here it is important to emphasize that it does not matter what the phase-one universal screener is measuring. *All that matters is that performance on that universal screener is highly reliable and strongly correlated with performance on the phase-two assessment(s).* At Phase Two we want to make sure that we are measuring the skills and dispositions necessary to do well in the program service for which we are identifying students (Lee et al., 2020; Lakin, 2018). But Phase One is not about identification; it is simply about deciding who is most likely to succeed on the identification (Phase Two) assessment(s). This is why we can focus on high reliability instead of measuring the "right" thing in Phase One. In the following section, we will describe a school district's current practice that uses an effective two-phase system for identification.

Sandra Day O'Connor School District

An example might make all of this clearer. The Sandra Day O'Connor School District identifies students for its highly gifted program using the *Otis Lennon School Ability Test* (OLSAT; Maddux, 2010), the Measure of Academic Progress (MAP; Northwest Evaluation Association, 2019) achievement test, and the Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003). In Phase One, district educators universally screen students, based on MAP scores for students who are scoring at or above the 90th percentile. Then in Phase Two, the three data points are put on the same scale (e.g., standard or index scores) and averaged together. Students must score at the 95th percentile (the composite average) on these data points to be eligible for the gifted program. To estimate the sensitivity of the O'Conner School District system, we will need the measures' reliabilities, intercorrelations, and the designated cut-scores. From the OLSAT (Maddux, 2010), MAP (Northwest Evaluation Association, 2019), and GRS (Pfeiffer & Jarosewich, 2003) technical manuals, we find reliability levels of .90, .97, and .97 respectively. Let us pretend that the correlation between the MAP and OLSAT is .70 (based on experience with similar assessments), and we estimate the GRS is related to both the OLSAT and MAP at something like .40. We assume a cut score at the 90th percentile, meaning that a student would only have to be in the top 10% of students to be identified. Averaging these three data points at Phase Two increases the reliability to .97 (see McBee et al., 2014 for further explanation). Finally, using the data points described, we assume the nomination validity of .86, meaning that the skills and abilities required to meet the criteria of the MAP screener are similar skills and abilities to do well on the GRS and OLSAT. The result of this estimated system is a sensitivity of .72. This system does a pretty good job of balancing cost (only the top 10% of students go through Phase Two) with sensitivity.

What does all of this mean? Effective two-phase systems are possible, but require consideration of the reliability of the screener, the cut-scores, and intercorrelations among the assessments. To reiterate, sensitivity would decrease if selective screening was used or if the nomination cut score was raised. A referral-based system that requires a certain level of performance on a rating scale and/or checklist from parents or teachers that was less related to the Phase Two assessments, would result in fewer students qualifying for formal evaluation. This was not the case at O'Connor School District; they used a screener that is highly related to Phase Two and had a lower nomination cut-score to the 90th percentile, resulting in greater sensitivity (.72), inclusive of more students from traditionally underrepresented groups. Although there would be an increase in cost to test more students in Phase Two, the added benefit is that more qualifying students could be appropriately placed in gifted services.

How to Improve the Quality of a Multi-Phase Identification System

Universal screening works to improve the sensitivity and, consequently, the equity of identification systems by improving (a) nomination validity, (b) reliability, and (c) removing the kinds of subjective screening or referral criteria that have been shown to miss students from traditionally underrepresented groups (Card & Giuliano, 2016). But universal screening systems are also less sensitive than universal consideration systems. Essentially, the challenge is whether an identification system can be crafted with the efficiency benefits of a well-designed two-phase system, but with the sensitivity of a single-phase system. The simple answer is no, but it is possible to get close. Recall from above that the single-phase, universal consideration system came with a sensitivity of .82, but at 100% of the cost since all students were being tested. Alternatively, the two-phase system came with a sensitivity of .51 (62% of the sensitivity for a universal consideration system [.51/.82]) and 10% of the cost since only those who scored in the top 10% of the phase-one assessment were assessed at Phase Two. There are two ways to improve the sensitivity of the two-phase system: (a) lower the phase-one cut score or (b) increase the nomination validity. A third option is to increase the overall reliability of scores in the identification system, but given we already assumed a very high-reliability score (r_{xx} = .95), this is an impractical solution.

The easiest way to improve the sensitivity and increase equitable access in two-phase identification systems is to lower the phase-one cut score and allow a larger number of students to be assessed in Phase Two. Too many districts require high cut scores for formal identification (e.g., 97th percentile in Arizona), but then also require a high cut score on their universal screener to even be tested (e.g., the 95th to 97th percentile). This is where many students are missed by the system. Students who would have scored at the 95th percentile in Phase Two, and thus been identified, but only scored in the 89th percentile or lower on the universal screener are overlooked. Often students from underrepresented groups are the ones who score lower, not necessarily due to a lack of talent, but instead due to larger societal inequality (Peters, 2021). The solution, therefore, is rather obvious—lower the phase-one cut score. If that score was lowered to the 80th percentile from the 90th, the sensitivity would increase from .51 to .67, meaning there would be an increase in sensitivity for limited cost (i.e., only 20% of the cost of a universal consideration system).

If the phase-one cut score was lowered even further to the 70th percentile, the sensitivity would increase to .74 for 30% of the cost. To be clear, such a system would still miss more students than would a universal consideration system, but it would also cost less (30%) than a universal consideration system. This exemplifies why there is always a trade-off between efficiency (i.e., cost, time) and effectiveness in finding underrepresented groups (i.e., sensitivity) when constructing an identification system.

Alternatively, a second way to improve the sensitivity of a two-phase system is to improve the nomination validity. In the Sandra O'Connor School District example, we assumed the correlation between the phases was .86, but if the nomination validity was lower (.60), we would see a significant loss of sensitivity; meaning more students would be incorrectly missed. As seen in the example, if the nomination validity is high (or increased), there can be greater sensitivity without lowering the phase-one cut score and spending more time and money on tests. Unfortunately, this is the harder of the two ways to increase sensitivity because in most circumstances it would involve purchasing and administering a different universal screening assessment. Perhaps, the existing achievement test is administered for other purposes and, as such, is not able to be discarded in favor of a different achievement test. Alternatively, the district could purchase and administer a new universal screener to all students in a given grade level, but the improved nomination validity would come at the cost of testing 100% of students. Odds are this is a higher cost than simply lowering the phase-one cut score as previously described.

Criteria for Universal Screening Instruments

Thus far, we have discussed reliability, nomination validity, efficiency (cost, time), and sensitivity in relation to universal screening and discussed how a specific two-phase system could be effective and improved. However, we have intentionally spent little attention on types of criteria because it is largely dependent on the usage of the instrument in the process. As long as the instruments used for universal screening measure what they are intending to measure, have evidence of performance consistency, are strongly related to Phase Two, and align with what students will encounter within the specified gifted program, any criteria can be used for testing (Crocker & Algina, 1986; Peters et al., 2020).

Quantitative versus Qualitative Criteria

The criteria for universal screening can be either quantitative or qualitative. Quantitative criteria provide numerical, quantifiable data to be analyzed. Quantitative criteria include cognitive ability tests (e.g., Cognitive Abilities Test [CogAT], Lohman, 2012; Naglieri Nonverbal Ability Test [NNAT], Naglieri, 201; Otis Lennon School Ability Test [OLSAT], Maddux & Morse, 2010; Screening Assessment for Gifted Elementary and Middle School Students [SAGES], Johnsen & Corn, 2018), achievement tests (e.g., MAP, Northwest Evaluation Association, 2019; Iowa Test of Basic Skills [ITBS], Hoover et al., 2001), norm-referenced creativity or divergent thinking tests (e.g., Torrance Test of Creative Thinking, Torrance, 2008), and other standardized tests that provide numerical data points. Standardized assessments can minimize subjective bias in the identification process and increase representation if implemented appropriately (Card & Guiliano, 2016; McBee et al., 2016). Similar to gifted identification, post-secondary institutions often require standardized test scores in their admissions process. Hyman (2017) found that when Michigan implemented a universal policy requiring all public school students to take a college entrance exam (i.e., ACT), high-poverty areas had significantly higher percent enrollment in a four-year institution; thus, exemplifying how standardized tests can help students from lower socioeconomic backgrounds.

Conversely, qualitative criteria included in the admissions process, especially selectively collected, can be problematic. For instance, Alvero et al. (2021), in a study of 60,000 undergraduate applicants, found that admissions essays were more strongly related to socioeconomic status (e.g., household income) than standardized test scores (i.e., SAT). Thus, the usage of qualitative criteria, such as admissions essays or writing samples, can inject more class bias into the identification process than would more traditional quantitative criteria (e.g., standardized ability or achievement tests mentioned previously).

Although much of what we have discussed is the value of quantitative criteria as universal screeners, qualitative criteria can be used too. The premise of using qualitative criteria is to provide a subjective assessment of a student beyond what is measured in a standardized measure; however, it is still often quantified for inclusion in the identification process (as seen in Sandra Day O'Connor School District). Popular qualitative criteria used for universal screening include gifted checklists and rating scales (e.g., Gifted Rating Scales [GRS], Pfeiffer & Jarosewich, 2003; Scales for Identifying Gifted Students [SIGS], Ryser and McConnell, 2004), and the Scales for Rating the Behavioral Characteristics of Superior Students [SRBCSS], Renzulli et al., 2010) that have raters (e.g., parents, teachers, self) evaluate specific characteristics associated with giftedness. Rating scales are commonplace in universal screening, but we urge caution in their usage since the reliability of scores can greatly vary. They can have weaker relationships with what is assessed at Phase Two, and inject more bias (i.e., error, unfairness) into the screening process. This is especially the case when districts decide to create their own qualitative checklists that may have weak validity.

We want to reiterate that qualitative screeners used in a referral-based two-phase system have shown to have weaker nomination validity, thus can harm sensitivity and ultimately nominate less traditionally underrepresented students that could be appropriately placed.

More sparsely used, alternative qualitative criteria include the use of performance-based task assessments (e.g., constructing a model, writing tasks), interviews, observations, or specific work samples (e.g., portfolios, writing samples) collected from all students assessed on a designated rubric. It is uncommon to use these specific qualitative criteria as a universal screener due to the time-consuming nature of collection and scoring; however, some schools collect work samples as an ongoing screening system (e.g., watchlists) for gifted programs (Mun et al., 2021). Additionally, districts may find value in using multiple pathways to formal evaluation with the use of multiple screeners (i.e., using both qualitative or quantitative at Phase One) to universally screen students for formal evaluation despite the increase in cost and time to implement. As noted previously, whether the criteria are quantitative or qualitative, a good screener possesses the same qualities—high reliability and strong relation to what will be measured at Phase 2.

Does the Test Matter?

Although the test is a fundamental part of the screening process, there is often too great a focus on which specific instrument to use, rather than how to use it (Carmen et al., 2018; Peters et al., 2020). Both quantitative and qualitative assessments can be used as screening tools. For example, nonverbal ability tests (e.g., Naglieri Nonverbal Ability Test [NNAT], Naglieri, 2018; Cognitive Abilities Test [CogAT] Nonverbal Battery, Lohman, 2012) have become popularized for universal screening due to the perception that they provide additional sensitivity for culturally, linguistically, and ethnically diverse (CLED) students. However, simply selecting a nonverbal test does not solve equity challenges within gifted education. For instance, Lohman et al. (2008) found that English Learners who took three different nonverbal tests (i.e., NNAT, Raven's Progressive Matrices, CogAT Nonverbal Battery) scored .50 to .60 standard deviations lower than native English speakers. Carman et al. (2020) found the NNAT2 and the nonverbal battery of the CogAT7 to be less likely to identify traditionally underrepresented demographics;

rather it was the collective decisions of the instrument and norming procedures that influenced identification.

Although the use of nonverbal ability tests is popular, other assessments can be used for screening that may identify more students for appropriate learning opportunities and better align with the gifted program curriculum outcomes (e.g., ability tests, achievement tests, performance-based creativity assessments; VanTassel-Baska, 2008). Which instrument to use at Phase Two largely depends on the skills and abilities that will be important to success in the resulting service.

Rather than a continued focus on the current tests used, we urge practitioners to consider the reliability of scores on assessments, the relation of the phase-one screener to phase-two, arbitrarily high cut scores, norming procedures within the identification process (Carman et al., 2020; Peters & Engerrand, 2016; Peters et al., 2019), and the alignment to the specific goals and outcomes of gifted programs (e.g., Lee et al., 2020; Gubbins et al., 2021; Johnsen, 2008). For instance, district personnel could focus on developing lower cut scores to the 60th–90th percentile to allow more students to be formally tested for gifted services rather than searching for a magic test to improve equity. Higher phase-one cut scores can be used, but that is highly dependent on how related the referral process is to scores within Phase Two (i.e., high nomination validity). The test itself matters as it relates to these phases and a gifted program's specific learner outcomes. Thus, district administrators need to be cognizant of issues of alignment of identification to gifted programs, as well as to this relationship between phases of the process.

Evaluating the Quality of an Identification System

There are three primary factors to consider when evaluating the quality of a gifted and talented student identification system: (a) sensitivity: the ability of the identification process to catch all of the students it was designed to find; (b) cost efficiency: the ability to keep testing costs low for materials and resources; and (c) equity: the ability to find representative groups of learners for gifted programs.

Sensitivity

The sensitivity of an identification process refers to its ability to correctly identify all of the students it was designed to identify (McBee et al., 2014). Importantly, sensitivity has nothing to do with an implicit values decision of which kids a school *should* be trying to find. This is based on a conceptual decision of giftedness that needs to be made first. Sensitivity involves how well the system is finding qualified students that have been designated as the target population of interest. If a school district has decided to identify the top 5% of students on a creativity test, then sensitivity means what percentage of those top 5% were correctly identified by the identification system that uses a creativity test.

There are several reasons why an identification system might miss kids. The simplest example is when a student was never put through the process because they were never referred to begin with. There are also more technical factors that can harm sensitivity, such as the use of subjective or otherwise less reliable assessment procedures (as discussed). If the less reliable scores are in an assessment process, then there will be worse sensitivity. Conditional on an operational definition of "gifted," sensitivity is how well a system correctly identifies students who meet a specified definition.

Cost

While not linear, when it comes to a quality identification system, on one hand, there is a relationship between sensitivity and equity and cost on the other. Balancing these factors is the topic of much of this chapter, but for now, we acknowledge a general rule: the more students who are assessed through the full identification process, the fewer will be missed (sensitivity increases). Similarly, the fewer students who are considered for services (such as only those who are referred), the more will be missed, but the process will be less expensive. Universal consideration is on one extreme where the most time and energy are devoted to identification with the outcome being the maximum level of sensitivity. The opposite is models that only consider those students who first meet several high criteria for consideration. In this case, testing is inexpensive, but many students who would benefit from gifted services go unserved. That said, the population of students missed will likely be disproportionately students from traditionally underrepresented groups (Card & Giuliano, 2016).

Equity

Identification systems can be equally effective for all student groups, or they can be more effective for certain subgroups than others. This unequal sensitivity is how we will conceptualize equity. This does not include all aspects of what should be considered when crafting an equitable identification system, but it is the primary factor when dealing with multi-phase identification procedures. An equitable identification system does not measure unrelated factors that are unique to particular subgroups. If a student must be nominated or referred before being considered for services, then the system will be biased toward students whose families are '*in the know*' meaning those from higher socio-economic backgrounds or have other forms of cultural capital who are comfortable advocating for their children (Walsh, 2008). Similarly, if a district allows families to bring in outside testing from a private psychologist, then those who have the means to pay for such services have a lower barrier to access gifted and talented services.

Equitable identification systems will not inherently result in perfect proportionality across student subgroups. An identification process can be bias-free and still result in the underrepresentation of students from certain traditionally underrepresented groups. This is because there is no such thing as a "pure" measure of ability of giftedness that is independent of educational opportunity and experience. As long as there is inequality of opportunity, there will be some inequality of outcome.

A Note on Universal Screening as an Equity Strategy

For clarity, universal screening is not a *solution* to inequity. Although universal screening, when done effectively, can help increase access to gifted services, other factors within identification systems can hinder equitable access. As discussed above, some students are missed by an identification system due to badly-designed two-phase identification systems. But that is not the sole or even primary driver of disproportionality in a gifted and talented population. Even systems of universal consideration will result in disproportionality and not achieve equal access to gifted programs. This is due to structural inequities that exist well before a student is ever tested for gifted programming (e.g., personal, economic, social, institutional; Peters, 2021). For instance, a student of color who lives in abject poverty, who also has limited English proficiency will have more obstacles in an identification process regardless of whether universal screening is implemented or not. There are broader societal inequalities that will continue to influence the differences that are found between demographic subgroups on any quantitative or qualitative assessment used in the identification process.

No system will ever be perfect, but consideration of how to improve the overall system and how to better use a screening tool (rather than the test itself) will reap the most benefits for students. Interested readers can consult Peters (2021) for a more in-depth discussion of this topic.

Recommendations for Coordinators

In this chapter, we have focused on identification systems because a systemic approach is needed to mitigate barriers to access in gifted identification whereby universal screening is a single component. Casting a wider net for students to be identified, as seen with universal screening and universal consideration, can promote equity. Improving sensitivity will almost always result in improved equity because making an identification system more accurate for everyone should by default make the system better for students from traditionally underrepresented groups (McBee et al., 2016). In addition to using universal screening or universal consideration, districts can use multiple screeners at various time points, acknowledge and change discrepancies in screening and formal identification processes, remove high thresholds for gifted service eligibility (i.e., cut scores), and implement data-driven leadership practices to evaluate their identification system (Mun et al., 2021). Major takeaways from this chapter on universal screening are as follows:

- Single-phase, universal consideration systems maximize sensitivity, but also maximize cost.
- Two-phase, universal screening systems, if carefully designed, can result in almost the same sensitivity as single-phase systems, but at far less cost.
- When choosing a universal screener, look for assessments that are aligned with programs, highly reliable, strongly correlated with phase-two criteria, and are as cheap and fast as possible.
- Set phase-one cut scores around the 60th–90th percentiles.
- Higher phase-one cut scores can be tolerated in the presence of high nomination validity.

References

Alvero, A. J., Giebel, S., Gebre-Medhin, B., Antonio, A. L., Stevens, M. L., & Domingue, B. W. (2021). *Essay content is strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications*. (CEPA Working Paper No.21-03). Stanford Center for Education Policy Analysis. http://cepa.stanford.edu/wp21-03

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences*, *113*(48), 13678–13683. https://doi.org/10.1073/pnas.1605043113

Carman, C. A., Walther, C. A. P., & Bartsch, R. A. (2018). Using the Cognitive Abilities Test (CogAT) 7 nonverbal battery to identify the gifted/talented: An investigation of demographic effects and norming plans. *Gifted Child Quarterly*, *62*(2), 193–209. https://doi.org/10.1177/0016986217752097

Carman, C. A., Walther, C. A. P., & Bartsch, R. A. (2020). Differences in using the Cognitive Abilities Test (CogAT) 7 nonverbal battery versus the Naglieri Nonverbal Ability Test 2 (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, *64*(3), 171–191. https://doi.org/10.1177/0016986220921164

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.

Grissom, J. A., Redding, C., & Bleiberg, J. F. (2019). Money over merit? Socioeconomic gaps in receipt of gifted services. *Harvard Education Review*, *89*(3), 337–369. https://doi.org/10.17763/1943-5045-89.3.337

Gubbins, E. J., Siegle, D., Ottone-Cross, K., McCoach, D. B., Langley, S. D., Callahan, C. M., Brodersen, A.V., & Caughey, M. (2021). Identifying and serving gifted and talented students: Are identification and services connected? *Gifted Child Quarterly*, *65*(2), 115–131. https://doi.org/10.1177/0016986220988308

Heinemann. (n.d.). Field study of reliability and validity of the Fountas & Pinnell Benchmark Assessment Systems 1 & 2. www.fountasandpinnell.com/shared/resources/FP_BAS_Research_Field-Study-Full-Report.pdf

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Tests of Basic Skills, Forms A, B, and C*. Riverside.

Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, *12*(3), 281–311. https://doi.org/10.1162/EDFP_a_00206

Johnsen, S. K. (2008). Identifying gifted and talented learners. In F. A. Karnes & K. R. Stephens (Eds.), *Achieving excellence: Educating the gifted and talented* (pp. 135–153). Pearson Education.

Johnsen, S. K., & Corn, A. L. (2018). *Screening assessment for gifted elementary and middle school students (SAGES-3) examiner manual*. Pro-Ed.

Lakin, J. M. (2018). Making the cut in gifted selection: Score combination rules and their impact on program diversity. *Gifted Child Quarterly*, *62*(2), 210–219. https://doi.org/10.1177/0016986217752099

Lakin, J. (2021). Service-identification alignment: Why the services you offer must drive identification decisions. [Manuscript submitted for publication]. *Teaching for High Potential*. https://drive.google.com/file/d/17qoHlyrX_8_GZAow9EqFm4hweWromqQI/view

Lee, L. E., Ottwein, J. K., & Peters, S. J. (2020). Eight universal truths of identifying students for advanced academic interventions. In J. H. Robins, J. L. Jolly, F. A. Karnes, & S. M. Bean (Eds.), *Methods and materials for teaching the gifted* (5th ed., pp. 61–80). Prufrock Press.

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 275–296. https://doi.org/10.1177/0016986208321808

Lohman, D. F. (2012). *Cognitive Abilities Test Form 7: Research and development guide*. Riverside.

Maddux, C. D. (2010). [Test review of Otis-Lennon School Ability Test, Eighth Edition]. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The Eighteenth Mental Measurements Yearbook*. Buros Center for Testing.

Maddux, C. D., & Morse, D.T. (2010). Otis-Lennon School Ability Test, Eighth Edition. In R.A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 875–879). Buros Institute of Mental Measurements.

McBee, M.T. (2016). What you don't look for, you won't find: A commentary on Card and Giuliano's examination of universal screening. *Journal of Advanced Academics*, *27*(2), 131–138. https://doi.org/10.1177/1932202X16634141

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69–89. https://doi.org/10.1177/0016986213513794

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

Mun, R. U., Ezzani, M. D., Lee, L. E., & Ottwein, J. K. (2021). Building systemic capacity to improve identification and services in gifted education: A case study of one district. *Gifted Child Quarterly*, *65*(2), 132–152. https://doi.org/10.1177/0016986220967376

Naglieri, J. (2018). *NNAT3 manual: Levels A-D.* Pearson. www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/nnat3/nnat3-manual-levels-a-d.pdf

Northwest Evaluation Association (NWEA). (March, 2019). *Technical manual for measures of academic progress (MAP).* Northwest Evaluation Association.

Peters, S. J., & Borland, J. H. (2020). Advanced academics: A model for gifted education without gifted students. In T. L. Cross & P. Olszewski-Kubilius (Eds.) *Conceptual frameworks for giftedness and talent development* (pp. 289–316), Prufrock Academic Press.

Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, *60*(3), 159–171. https://doi.org/10.1177/0016986216643165

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who gets served in gifted education? Demographic proportionality and a call for action. *Gifted Child Quarterly*, *63*(3), 273–287. https://doi.org/10.1177/0016986219833738

Peters, S. J., Ottwein, J. K., Lee, L. E., & Matthews, M. S. (2020). *Identification.* In J. A. Plucker & C. M. Callahan (Eds.), Critical issues and practices in gifted education (pp. 261–272). Prufrock Press.

Peters, S. J. (2021). The challenges of achieving equity within public school gifted and talented programs. *Gifted Child Quarterly*. https://doi.org/10.1177/00169862211002535

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales: Technical manual.* The Psychological Corporation.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., Gavin, M. K., Reis, S. M., Siegle, D., & Reed, R. E. S. (2010). *Scales for Rating the Behavioral Characteristics of Superior Students: Renzulli Scales*. Prufrock Press.

Romer, N., von der Embse, N., Eklund, K., Kilgus, S., Perales, K., Splett, J. W., Sudlo, S., & Wheeler, D., (2020). *Best practices in social, emotional, and behavioral screening: An implementation guide. Version 2.0.* https://smhcollaborative.org/universalscreening

Ryser, G. R., & McConnell, K. (2004). *Scales for Identifying Gifted Students*. Prufrock Press.

Torrance, E. P. (1990). *Torrance Tests of Creative Thinking—Technical Manual Figural*. Scholastic Testing Service.

Torrance, E. P. (2008). *Torrance Tests of Creative Thinking: Norms-technical manual figural (Streamlined) forms A & B*. Scholastic Testing Service.

Van-Tassel Baska, J. L. (Ed.) (2008). *Alternative assessments with gifted and talented students. The critical issues in equity and excellence in gifted education series*. Prufrock Press.

Walsh, P. (2008). Are involved parents providing public goods or private goods? *Public Finance Review*, *36*(6), 678–705. https://doi.org/10.1177/1091142107308298

CHAPTER

3

Quantitative Assessment Tools for Identification

Ability and Achievement Tests

Joni Lakin, Monica Simonds, and Marla Caviness-French

Introduction

Assessments that yield quantitative data are the cornerstone of many identification processes for gifted and talented programs or advanced academic services. The data from these assessments provide context about student potential alongside artifacts of student work, educator judgments of academic promise, and observational data.

Tests are only one tool and, when used well by professionals, can provide important information. In this chapter, we focus on the core tools and knowledge that educators need in order to make effective use of quantitative assessment data for identifying gifted learners. We begin with general theories and principles that apply to all assessments and then describe the specific features and uses of achievement and ability tests.

Student Characteristics Related to Academic Success

Across the identification process, the goal should always be to match student needs to educational services and strategies to maximize student learning (Lakin, 2018b). The needs of students and the types of services that are available to maximize learning should precede any decisions about what assessment data are needed.

The choice of assessment tools should relate to the domains in which services will be offered. In thinking about identifying *potential* for advanced academic performance, Lohman (2005; see also Lohman, 1995) outlined the most important predictors of a student's future performance in an academic domain:

- their current performance in that domain
- their ability to reason in the symbol system of that domain
- their interest and motivation for that domain
- their ability to persist in challenging learning environments

Measures of current performance include achievement tests, classroom assessments or evaluations, and portfolios of student artifacts (Matthews & Peters, 2018; Ryser,

DOI: 10.4324/9781003285991-5

2018). Measures of reasoning abilities usually take the form of cognitive ability measures and, sometimes, observations by teachers and parents. In this chapter, we will focus on assessments of achievement and ability as well as the quantitative scores that they produce. However, other characteristics, including interest and motivation, are vital to learning and are an important part of identification and aligning services to students capable of advanced learning (NAGC Standards 1.5, 3.2, 2.2). Supporting learners' success requires measures of each of these characteristics to build from areas of strength and to address areas of need to ensure success.

The Structure of Human Abilities

The Cattell-Horn-Carroll (CHC) model of intelligence is by far the most widely-accepted and empirically-supported model of human ability (McGrew & Schneider, 2018). At its simplest level, this model maintains that students have different levels of various skills that contribute to learning and everyday problem solving, but that these skills are correlated—i.e., students will tend to have generally high or generally low scores across the broad abilities, although there can be some peaks and valleys in performance. This observation of correlations between disparate abilities is called *g*, general ability, or general intelligence (Cooper, 2015; Deary, 2001).

Under *g* is a small set of broad abilities that are a little more specific in the contexts where they are demonstrated. The most important of these include crystallized intelligence (*Gc*) and fluid reasoning (*Gf*). Crystallized intelligence reflects acquired knowledge—what a person knows and can do based on formal or informal learning opportunities. In education, *Gc* is associated with achievement tests that are closely tied to curriculum and instruction. Fluid reasoning, on the other hand, is less tied to instruction and represents a more general ability to reason with ideas and to acquire new knowledge. Fluid reasoning is most associated with ability or aptitude tests. (Note that the terms ability and aptitude are often used interchangeably and do not have a meaningful distinction in modern assessments.) As two related aspects of human ability, measures of *Gf* or cognitive ability and *Gc* or achievement fall along a continuum, rather than into clear categories of assessments. See Figure 3.1 which shows a variety of common tests on a continuum from measures of fluid reasoning to crystallized knowledge or achievement (Lohman, 2006).

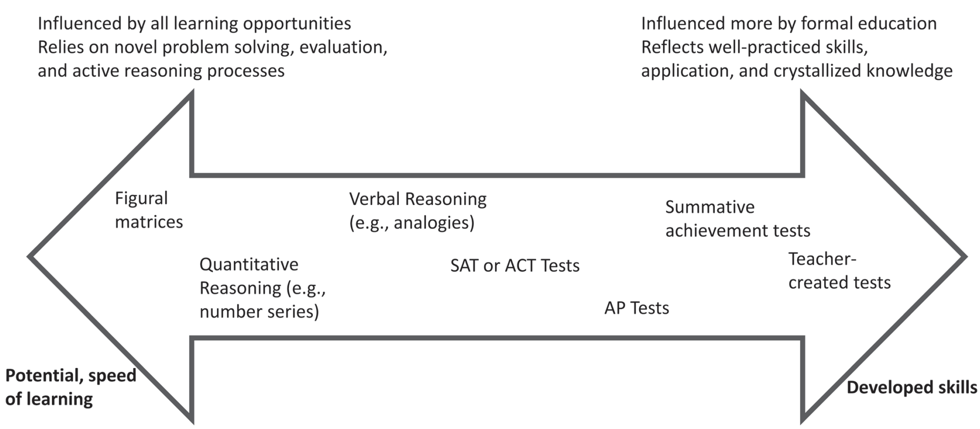

FIGURE 3.1 Differences in measures of ability and achievement arranged on a continuum from those that better measure potential to those that better measure developed skills

Alongside these two most important broad abilities are other broadly useful cognitive attributes including working memory. Below each of these broad abilities are an array of specific or narrow abilities, which represent even narrower ranges of cognitive skills that are useful in more limited contexts.

In the *2018–2019 State of the States in Gifted Education Report* (Rinn et al., 2020), 29 of 44 (66%) responding states indicated that "academic ability" was part of their state definition of giftedness and 36 (82%) reported that "advanced intellectual ability" was part of the state definition (p. 17). Based on the CHC model of human abilities, high levels of academic performance (i.e., high achievement scores) can be one indicator of high academic or intellectual ability because it reflects crystallized (*Gc*) knowledge, but this definition also encompasses measures that fall at the fluid (*Gf*) end of the continuum.

There are some important features of measures of this model that are useful to understand as an administrator. First, the most generally useful ability is termed general intelligence (*g*) and is defined not so much as a specific type of thinking, but as the level of typical performance across a large number of diverse types of cognition. In other words, people who are good at many different types of problem solving and types of thinking skills have high general ability. Thus, a second implication is that, in order to measure broad abilities or general ability, tests need to include a variety of subjects and reasoning tasks in order to average out the specific skills of a given test and provide a measure of general skills.

Finally, general ability is not the only important characteristic. Broad abilities can provide additional information for planning instruction around students' relative strengths. Variation in broad abilities, as well as the importance of interest and motivation, is part of the reason one-size-fits-all gifted education programming is ineffective.

Key Measurement Concepts

Assessments of ability and achievement yield a variety of scores with different interpretations or purposes. These will be useful in understanding the descriptions of assessments that follow. It is also important to understand what is meant by measurement error and measurement bias. We will explore these concepts in this section.

Norm- vs. Criterion-Referenced Scores

Scores supporting normative interpretations are based on comparing students to a large sample of students at the same age or grade level. Scores are often reported in percentile ranks, index scores, or other standardized scales and reflect how well the student performs in that skill set compared to similar students.

Criterion-referenced interpretations translate test performance into descriptive categories. For example, students may be sorted into categories of proficiency or "college and career ready." Students may also be classified as achieving or not yet achieving specific standards. These criteria are defined by standard-setting panels, composed primarily of educators who determine, through a systematic standard-setting process, the test cut scores that define criterion categories (Koretz, 2009). Thus criterion-referenced tests use standards as the benchmark for judgment while norm-referenced tests use comparable student performance as the basis.

The Special Case of Local Norms

Local norms involve comparing students not to a national sample of peers, but to just the students tested in their district or school building. These are almost always calculated as percentile ranks. Thus, you can identify the top 5% (or any percentile) of students locally (in other words the 95th local percentile) rather than using national norms that tell you which students are in the top 5% across the U.S. Many testing companies will provide local norms, based on annual testing or a district may calculate such norms themselves (Lakin, 2020).

Why are local norms useful? Gifted and talented services are most effective when they serve students for whom the standard curriculum in their school building is not appropriate. Schools can recognize these students based on local norms, which will more clearly reflect how much of the knowledge and skills the student has acquired compared to other students receiving the typical curriculum in a given school district. In other words, if a school identifies 40% of the population as gifted, they will have many students receiving services that may be well-served by the standard curriculum. If a school identifies 0–5% of students as gifted, then they will have a small percentage of students in the general education classroom who are bored and underchallenged and who need faster paced and deeper instructional content than the standard curriculum. Whether national norms identify too many or too few students in a district, local norms allow the school to manage the size of their gifted services population and appropriately tailor instruction for their local students (Peters et al., 2021).

Occasionally, educators will propose local norms as a strategy for increasing the diversity of students identified for gifted services. Sometimes this claim is based on confusing local norms with sub-group norms, discussed in the context of ability scores, where sub-group norms can help control for opportunity to learn. Local norms maintain the rank-order of students on any score scale, so they do not address differences in opportunity directly.

One scenario where the use of local norms may promote diversity is when there are marked demographic differences across school buildings within a district that leads to substantially different ability score distributions across schools. This would happen if some school buildings in a district have higher rates of poverty, for example. In this case, estimating local norms within school buildings will allow each school to identify students most likely to benefit from specialized services within that student population at that site. This will avoid a situation where some school buildings have "no gifted students," when the reality is that every school has students who require more challenge and accelerated learning than their school peers and more than the standard curriculum will offer.

Common Scores

Tests yield different types of scores. Because of the importance in their use and interpretation, we will discuss three of these: raw scores, index scale scores, and percentile ranks. **Raw scores** are usually just the number of test questions answered correctly or the points awarded from a rubric. Raw scores are never useful for identification purposes because they are not comparable across tests, batteries/domains, or forms.

Index scale scores, sometimes called standard or standardized scales, are mathematical transformations of raw scores that support comparability of scores for different purposes. Vertical scales are one kind of index score that link student performance across test *levels*, allowing users to compare test scores across sets of items designed for various grade or age-levels. Other index scales are designed to promote interpretability across *forms* of the test (i.e., different versions of a test built to the same difficulty) or different content areas or batteries of a test. Non-vertical standard scores are usually based on a normal distribution, termed normalized standard scores. The actual numbers can be virtually any range of numbers chosen by the test creator. A common standard scale is the IQ scale with a mean of 100 (*M* = 100) and standard deviation of 15 (*SD* = 15).

Percentile ranks (PR) express a student's test performance relative to a norm group. The norm group referenced should be a large, representative sample of similar students. Normative groups can be based on same-age or same-grade peers.

Score Use

For gifted identification, raw scores (number correct) should never be used as they are not comparable across test forms (versions), levels, or subject areas. Percentile ranks may be used for decisions based rank ordering or developing cut scores. However, index scores are often the most appropriate because they provide more precision at the highest levels of performance (assuming it is a test designed to distinguish exceptional levels of performance). Gifted and talented students often score at the 98–99th percentile for their age and grade, which means that many students are lumped together with the same score. If a school needs to differentiate students within this group, index scores will often provide more precision. For example, on level 8 of the CogAT, students with a percentile rank of 99% can range in the vertical Standard Age Score scale (i.e., index score) from 137 to 157—a wide range of underlying ability lumped together into one percentile score.

Measurement Error and Reliability

Educators interpreting differences among test scores should keep in mind that not all differences in performance across dimensions are meaningful. *Measurement error* (the random variations in test performance due to time, test content, and other impacts) alone will result in test scores that vary for an individual over administrations even when the exact same test is administered. The amount of measurement error contained by a test score is reflected by *reliability* statistics as well as by the *Standard Error of Measurement* (*SEM*), another estimate of consistency based on the reliability statistic. Reliability statistics are often based on estimates of internal consistency (i.e., the degree to which students' performance on individual items correlate) or from a test-retest study, where students take the test twice to estimate the consistency of their scores over time (Robins & Jolly, 2018). According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014), all assessments should provide information about reliability in their technical materials.

Test reports often include the *SEM* or translate that value into a *confidence interval* (or score band) which reflects the range of scores that students might receive if they were hypothetically to take the test over and over. If confidence intervals are not provided, it is relatively simple to calculate. If the test manual provides the *SEM*, you can add and subtract one *SEM* from the student's test score to create a 68% confidence interval, which reflects moderate certainty (68%) that the interval contains the student's true score level. For higher precision, you can add and subtract about 2 *SEM*s (1.96 to be exact) to create a 95% confidence interval. If an *SEM* is not provided, it can be calculated using

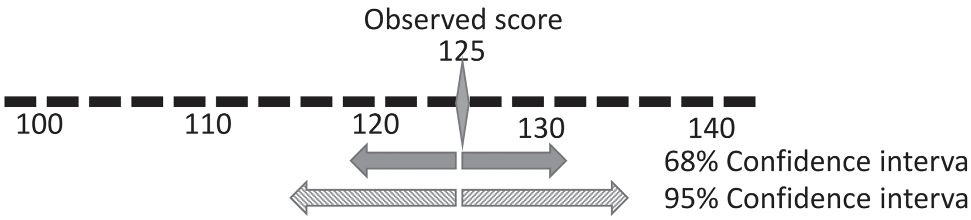

FIGURE 3.2 A student receiving a scale score of 125 could range in their true performance level, based on measurement error and reflected by the confidence interval

a reliability statistic or internal consistency estimate as well as the standard deviation of scores. The formula is $SEM = SD * \sqrt{1-r}$ where the reliability estimate (r) is subtracted from 1, then the square root is taken of the difference ($1-r$) and the square root difference is then multiplied by the standard deviation of scores (SD).

For example, if a student receives an index score of 125 on test A, that score contains some amount of measurement error which suggests the "true score" may be somewhat higher or lower than that score. In the example in Figure 3.2, the *SEM* for that test is 5, then we get the 68% confidence interval shown in Figure 3.2 (120–130). We would be fairly confident that this range includes the student's true score. If we want to be more cautious about concluding that one score is significantly different from another (such as making referrals for special education services), we might use the 95% confidence interval. In this case, we add and subtract 1.96*SEM (19.6) from the score to get a confidence interval of 115 to 135.

If we were making decisions using a cut score (say 130), we would see that the student's reported score did not meet the cut score, but that their range of *plausible* true scores does include 130. This student should receive additional consideration even though they did not meet the exact cutoff score.

If we administered another test, we would also want to use the confidence interval, *not* the actual scores, to compare and decide if one was notably higher or lower than the original score. If the confidence intervals do *not* overlap, the scores are significantly different. If they do overlap, we would conclude the differences may be due to measurement error and not a meaningful difference in scores.

As a general rule, subscores from a test (such as math computation scores), which have a fewer number of items, will be less reliable than the composite score (such as math total score) and will have a higher SEM, which is reflected in the confidence interval. Higher stakes decisions, such as course placement or gifted identification, should be based on scores that are sufficiently reliable. This may mean using battery-level scores and not narrow skill scores. On the other hand, individual skill scores are often more useful for planning instruction than the battery-level scores. For example, teachers will get more information by knowing a student has weaker spelling skills compared to their grammar skills than they would get from learning that a student's ELA scores are weaker than their mathematics scores. Therefore, it is important for gifted and talented coordinators to understand how to incorporate battery and subtest scores into the GT identification and instructional process.

Measurement Bias

Measurement error varies across time, student, and test questions and needs to be considered when interpreting test scores. In contrast, *measurement bias* is the observation that the test is measuring something other than what it is intended to measure (e.g., a student's proficiency with language rather than mathematics). A good metaphor is the digital scale many people have in their homes. Measurement error is the random fluctuations in weight observed due to minor variations throughout the day (e.g., being slightly more or less hydrated). On that same scale, measurement bias is when the scale consistently measures 5 pounds lower than the true weight. Group bias would be the situation if the scale only measured five pounds off for men but 10 pounds off for women. Measurement error is always present to some degree but does not cause bias. Systematic effects (bias), however, must be addressed.

Developers of standardized tests use many strategies to minimize bias and promote fairness in assessments. Again, the *Standards* (AERA, 2014) guide professional test developers in the types of evidence they need to provide to demonstrate fairness and lack of measurement bias in assessments. Some of these tools and evidence include fairness and sensitivity reviews, where educators and other stakeholders review test items for fairness and lack of problematic content. Another type of evidence is *differential item functioning* (*DIF*), which detects when individual items are substantially harder or easier for specific groups of students.

Many educators believe that tests have been shown categorically to be biased against marginalized groups in the U.S. educational system. Although there is always room to improve methods (Jonson et al., 2019; Zwick, 2019), tests of all kinds are carefully screened and revised to minimize the cultural content and to maximize fairness (AERA et al., 2014; NAGC Standard 2.3). Bias has not been demonstrated consistently against marginalized groups in standardized tests used to identify gifted populations.

Opportunity Gaps

In contrast, educational opportunity and access to quality pre-K is not available in many locales and differs substantially across schools and communities where it is available (i.e., opportunity is distributed in distinctly biased patterns, Condron, 2009; Fryer & Levitt, 2004). These differences in opportunity are *reflected* in test scores. Tests do not cause them. As the National Council on Measurement in Education (NCME) stated, "criticizing test results for reflecting these inequities is like blaming a thermometer for global warming." (p.1, NCME, 2019). This is why many equity researchers call for thinking of achievement "gaps" as educational "debts" to students who have historically not received equal access to quality education (Ladson-Billings, 2006). Both achievement and ability tests *reflect* systemic inequalities in society and can actually be a force for documenting and addressing these gaps (Erwin & Worrell, 2012; Sireci, 2021).

Achievement

Definition

Achievement tests measure developed knowledge and skills that are directly tied to the school curriculum, which in turn is calibrated to the standards. The results of an achievement test indicate the extent to which an individual has mastered a body of knowledge or a set of skills. Most achievement tests break down scores into specific domains or skill sets to provide more detailed information about which skills students have acquired and which need more focus, such as mathematics tests which may report separate scores for computation, algebraic thinking, and use of data.

Purpose and Characteristics

Important distinctions among achievement tests are those designed for *formative* and *summative* purposes (Koretz, 2009). Formative (also termed *interim*) assessments are usually closely tied to the curriculum and intended to be administered relatively frequently, at least several times per year, to assess students' gains in knowledge and to plan instruction in the short-term. Because these tests are used for *low stakes* decisions (meaning that no long-lasting changes for students or teachers are made using the results), these assessments are often intentionally brief and adapted to reflect student skill levels. Summative achievement tests are used to assess learning at the end of a unit of instruction or a significant block of time (e.g., end of grade). Summative tests usually have higher stakes attached to them, such as grade promotion, course placement, or teacher evaluations. Therefore, summative tests tend to be longer and intended to yield more reliable scores. They may be designed based on a state's curriculum or aligned to a more general set of standards. The main purpose of summative assessments is to provide an evaluation of students' mastery of grade-level skills at a more global level. These assessments also often provide information about specific skills but tend not to be as sensitive to short-term instructional changes as are formative assessments.

Achievement tests often report both normative and criterion-referenced score interpretations. Both normative and criterion-referenced interpretations for achievement tests are most often based on grade-level comparisons. Normative interpretations include grade percentile ranks and index scores.

Criterion-referenced interpretations most often translate achievement test performance into categories based on grade-level expectations found in the curriculum standards for a given state. For identifying, placing, or evaluating students for gifted education services, criterion-referenced interpretations around *current* grade-level proficiency are less useful than norm-referenced scores. As a result of how the criteria are defined, students eligible for GT services will tend to be lumped together in the proficient and advanced proficient categories, which gives us limited information about their actual skill levels or instructional needs. Even *career and college ready* standards (Conley, 2014) are usually focused on a lower level of achievement than the performance of many gifted students.

Special Cases of Achievement Tests

Above-level testing involves administering a test designed for older students (at least two grade levels above) so that the content better matches the level of challenge some students are capable of and allows students to demonstrate content knowledge and skills well beyond what is typical for their grade level (Rambo-Hernandez & Warne, 2015). Students who learn quickly or who have explored a topic independently can often perform well above their current grade level skills and content, reaching the upper limit of an achievement test, referred to as the "ceiling effect." In other words, the test items are not challenging enough to distinguish between the performance level of students who are advanced learners, who all receive similar scores despite substantial variation in their actual skills.

Above-level tests are more commonly offered through national *talent search programs*, such as the Center for Talent Development at Northwestern, where, for example, seventh and eighth grade students take tests like the SAT or ACT, which are designed for juniors and seniors in high school. Using above-level tests for younger students results in scores that are closer to ability or reasoning skills (see the continuum in Figure 3.1). Knowing that students can achieve much beyond grade level can inform not only decisions about identification but also a student's potential in a specific domain and teachers' decisions about the needs of talented students. This may include curriculum compacting, grade-level acceleration and/or subject acceleration (e.g., dual enrollment options, Advanced Placement), or finding a mentor to guide the student's advanced knowledge and skills).

Another important consideration is whether the achievement test supports measuring student *growth*. With testing systems that involve repeated assessment of students, schools increasingly have access not just to information about the student's current performance, but also whether they have grown more or less than other students. Growth rates can also be an important indicator for identifying those students who need accelerated services. Students engaged in advanced academics should show growth in skills over time (NAGC [2019] Standard 2.4, 5.2). Either formative or summative assessments may provide estimates of student growth, but teachers should not assume that all assessments provide reliable estimates of growth, especially for students testing near the score ceiling. Changes in test content over time easily invalidate the growth interpretations, so only growth estimates supported by strong validity evidence should be used (Briggs & Peck, 2015).

Challenges and Misconceptions

It is a myth that students with exceptional academic ability will excel in all domains of academic achievement. As we will discuss in the context of ability scores, it is quite common for students with exceptional abilities in one area to have relative areas of weakness. These differences can have many different causes, such as educational opportunity, learning disabilities, lack of interest, or predisposition.

Importantly, longitudinal studies of students that consider these profiles of scores demonstrate that not only do exceptionally talented students sometimes have areas of weakness, but these profiles predict the kinds of expertise that students ultimately develop (Lubinski & Benbow, 2006; Makel et al., 2016; Park et al., 2007; Wai et al., 2005). In other words, having areas of talent alongside weaker areas of performance are somewhat common in gifted students' profiles and are pertinent to planning appropriate services for students.

Motivation

Another misconception is that achievement tests are always accurate indicators of academic potential. Achievement tests are a product of cumulative learned educational experiences. If students are under-motivated in the classroom, their achievement scores may not reflect their full capability. Even though current achievement is important in predicting future achievement (Sternberg, 2011), students vary in their interest and motivation over time and across different content areas. Students also need to be appropriately motivated (but not made anxious) for performing well on assessments. Students who are bored or "checked out" can have artificially low achievement scores.

Twice-exceptionality

Students with both high academic ability but also specific learning disabilities are referred to as twice-exceptional (Foley-Nicpon et al., 2013). These students may be perceived as gifted students who are underachieving and need to make more effort or may appear as students served by special education who do not need academic challenge because they need other supports. Students with specific learning disabilities should be evaluated for gifted education needs in ways that are sensitive to their disabilities, such as using nonverbal tests for students with language-related disorders. Training teachers to look for evidence of need among students with specific learning disabilities can also help programs identify these students even when their achievement scores are not high enough to meet traditional identification standards.

Likewise, general education and gifted education teachers need to be aware of the potential for specific learning disabilities among gifted students. If students show remarkably low achievement in just one academic domain or struggle with a narrow range of academic skills (e.g., just spelling or just basic computation), this may be a clue that diagnostic testing is needed to identify learning difficulties.

Scope of Assessments

Another potential challenge is the scope of the content assessed, which may not include a student's area of strength or interest. Students who qualify for gifted services often have diverse and varied interests beyond the scope of the general education content. As they may learn more quickly than the content is presented and they may pursue areas of interest more deeply than the curriculum presents, achievement tests will not be an accurate measure of what the student actually knows. It simply provides an assessment of what a student knows and can do in relation to the test's scope based on the school's curriculum. Therefore, a teacher providing gifted services may need additional tests or student portfolios to plan instruction that can allow students to demonstrate growth in specialized domains.

Summary

Because achievement tests measure what a student has learned, if a student has had any obstacles to learning, their achievement scores will not reflect their *potential* to learn new material. Achievement scores may be less reflective of potential for students with English as an additional language, from under-resourced homes or schools, coming from an economically disadvantaged background, or being twice-exceptional. Therefore, if achievement tests are being used to determine gifted services, the assessment itself becomes a gatekeeper for these student groups, exacerbating the effects of an already inequitable system. One of several ways to assess *potential* in addition to current high achievement is to assess cognitive ability. We turn to these tests next.

Cognitive Ability Assessments

Alongside achievement tests, measures of cognitive ability are the most commonly used assessments in identifying students for gifted and talented services (NAGC, 2015). They provide a different perspective on students' academic readiness, reflecting more the *potential* for advanced learning.

Definition

Cognitive ability tests measure reasoning skills that have been acquired through all of a child's experiences, not just formal education. Specifically, cognitive ability tests call on skills such as problem solving, noticing patterns and connections, inferring meaning from incomplete information, and combining current knowledge in novel ways. Individually administered intelligence tests usually measure many of the broad abilities defined by the CHC model discussed earlier. Most group-administered cognitive ability tests, particularly those used for identification, measure fluid reasoning (*Gf*). As discussed earlier, fluid reasoning emphasizes a person's capacity for reasoning and learning from experience. They seek to reflect a student's ability to learn quickly or to abstract their learning more readily (i.e., finding patterns or logical connections).

Ability tests use a variety of question types that engage students in (ideally) novel problem-solving to demonstrate the skills above. Analogies (verbal, numerical, and figural) are common formats across many domains of reasoning. Specifically verbal ability tasks can include recognizing synonyms and antonyms, answering riddles, identifying categories, and filling in blanks to create a meaningful sentence. Quantitative tasks include number series (pattern recognition), comparing quantities, or word problems. Figural tasks often include analogies (usually called figural matrix formats), series and patterns, and classification tasks, such as finding the figure that completes a set or finding the figure that does not fit a set ("odd man out").

Purpose and Characteristics

There are a few categories to consider among the cognitive ability tests available. First, there are individually administered ability tests, such as the Wechsler Intelligence Scale for Children (5th ed. [WISC-V], 2014), and group-administered ability tests, such as the Cognitive Abilities Test (CogAT 8,Lohman & Lakin, 2017). Individually administered ability tests, which incorporate multiple subtests to administer and interpret, require extensive training while group-administered tests can generally be administered and interpreted by a teacher with minimal training. Individually administered ability tests can be used for clinical decisions such as diagnosing specific learning disabilities. Group-administered tests are better suited for referring students for further testing using individually administered tests. Group-administered tests are also more feasible and cost-effective for gifted and talented universal screening.

Another distinction with cognitive ability tests is whether they are unidimensional (i.e., provide one global score) or multidimensional (i.e., provide many different scores distinguished by content or domain such as verbal and quantitative reasoning. Among individually administered tests, "brief" forms are often unidimensional, resulting in just one primary score of general ability, while full assessments provide a number of test scores with different interpretations or uses. For example, the WISC-V reports Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed composite scores as well as the Full Scale score as part of its core battery. Each composite score is based on at least two subtests in a particular format, which create an average score for a specific ability (Conway et al., 2021; Süß & Beauducel, 2005). Table 3.1 provides examples of both individually and group-administered tests that are unidimensional and multidimensional.

How does IQ relate to cognitive ability in the CHC model? IQ (which stands for intelligence quotient) is not part of the modern theory of human ability, although the term is still used as a synonym for general cognitive ability (Bartholomew, 2004). The IQ scale (M = 100, SD = 15) is used as a familiar index scale for normally distributed ability test scores, but the original interpretation (a ratio of mental age to chronological age) is no longer used.

TABLE 3.1 Examples of multi- and unidimensional ability tests

Format	Dimensionality	Example assessments	Domain scores beyond general ability
Individually administered, requires advanced training	Unidimensional	Wechsler Nonverbal (WNV)	N/A
		Comprehensive Test of Nonverbal Intelligence (TONI-4)	N/A
	Multidimensional	Woodcock-Johnson (WJ-IV)	Comprehension-knowledge (Gc), Fluid reasoning (*Gf*), working memory (*Gwm*), Auditory processing (*Ga*), Long-term retrieval (*Glr*), Visual processing (*Gv*)
		Kaufman Assessment Battery for Children (KABC-2)	Simultaneous (*Gv*), Sequential (*Gsm*), Planning (*Gf*), Learning (*Glr*), and Knowledge (*Gc*)
Group-administered, minimal training	Unidimensional	Naglieri Nonverbal Ability Test (NNAT, 3rd ed.)	N/A
		Raven's Progressive Matrices (2nd ed.)	N/A
	Multidimensional	Cognitive Abilities Test (CogAT Form 8)	Verbal reasoning, Quantitative reasoning, Nonverbal reasoning
		Universal Nonverbal Intelligence Test (UNIT-2)	Memory, Reasoning, Quantitative

Interpreting Ability Scores in Culturally Diverse Contexts

The essential interpretation we are trying to make from cognitive ability tests is *how quickly or easily does this student solve problems and learn?* By using normative comparisons, specifically comparing students to others with similar opportunities to learn, we can interpret their ability test scores as an indicator of how easily they solve problems or how quickly they learn. As an example, imagine we gave two people some time to memorize as many made-up words as they could. Their performance on a test of these words would tell us who was most efficient at learning new words.

Obviously, with people in the real world, we do not know precisely what opportunities each student has to learn vocabulary, think about quantities, or solve puzzles with shapes and figures. Instead, we treat age as a proxy for opportunity to learn. In a perfectly equitable and homogenous society, age would be a good proxy for how much time each student had to learn language, math, and figures. In our imperfect society, age alone does not account for variations in opportunity, so adjustments need to be made to better account for opportunity to learn. When making interpretations about students' ability, we need better estimates of opportunity to learn than age-based norms can provide on their own. This is where local norms and subgroup norms can be valuable—to compare students to a group that is more similar to them in terms of opportunity to learn.

Local norms. As described earlier, local norms are based on comparing students only to other students at the same age or grade level *in their local district or even just their school building* (Peters et al., 2021; Yaluma & Tyner, 2018). Building local norms in particular can be useful for adjusting for differences in opportunity to learn when districts have larger variation in students' racial, ethnic, linguistic, or socioeconomic backgrounds across school buildings. However, when these characteristics are evenly distributed across schools, then local norms alone will not affect representation of students who are historically underrepresented (Warne & Larson, 2021; Worrell & Erwin, 2011).

Subgroup norms. Subgroup norms are a strategy that can result in greater identification of academic potential across marginalized groups who are underrepresented in the local gifted services. It is more effective for increasing diversity of identified students than local norms, particularly when buildings are similar in terms of demographic factors. Subgroup norms are also more appropriate, in many cases, than national age-based norms for correcting differences in opportunity to learn that uniquely impact the interpretation of ability scores.

Rather than basing student percentile ranks on the local population, subgroup norms involve creating ranks of students grouped based on background characteristics that systematically affect opportunity to learn. For example, the WISC-IV and V Spanish assessments offer *demographically adjusted percentiles*, which are sub-group norms based on variables like parent education or years in U.S. education, with a few categories of each created based on test information. These are very useful in identifying students from underserved groups for specialized services, because they allow test users to better compare students whose opportunities differ substantially and recognize those that learn most effectively and efficiently.

Carman (e.g., Carman et al., 2018) has an extensive research agenda on subgroup norms used in one large district where they essentially translate normative differences based on key groups (e.g., Latinx/Hispanic students, EL students, etc.) into *opportunity* or *modifier points*. They suggest applying these modifier points to scores in order to compare all students to the same cut-scores.

Another strategy that some districts and states use is to have more than one path to identification, using different assessments or cut scores. For example, Florida has a "Plan B" eligibility, where students who are English learners or who are eligible for free or reduced-price lunch (FRL), can meet a lower threshold for identification. District administrators, however, should be cautious in implementing modifiers or alternative paths to identification. In some states, these systems are illegal or disallowed when they result in essentially quota systems by race or other protected status.

Note that at the heart of both of these group-focused systems is a firm cut-score, which is sometimes unavoidable due to state regulations. When possible, holistic review, which include multiple assessments and expert judgment, will often lead to greater equity and sensitivity to student needs than assigning points or using alternative group norms under rigid cut-score rules (Erwin & Worrell, 2012).

Challenges and Misconceptions

Cognitive ability tests are vulnerable to a number of misinterpretations, most notably, the belief that they measure innate capacity and are not affected by opportunity to learn. The truth is that cognitive abilities are developed throughout the lifespan and are sensitive to educational advantages, especially in childhood where there can be a marked benefit from enrichment experiences.

A good metaphor is athletic skill. While there is certainly some genetic component to athletic ability, taking advantage of that predisposition requires a child to engage in exercise and specific training. At any point, we can measure a student's athleticism, but the assessment will reflect both innate skill and current level of development. We can never just measure the innate talent. Further, training leads to both specific gains (e.g., achievement in basketball) as well as general gains (e.g., increased physical strength) that can be a useful ability or readiness to succeed in other athletic domains. So, there is no physical strength without some kind of experience moving athletically and those who receive expert coaching may more fully realize their athletic potential. Likewise, although cognitive ability may have some genetic component, the abilities we are measuring in school are based extensively on learning opportunities and cannot be measured independently of experience or opportunity (Lohman, 2006).

Another myth about cognitive ability is that it only predicts narrow types of academic success. Life itself involves constant problem solving—such as figuring out how a new chip reader at the grocery store works—and researchers have found that general ability predicts both career and general life outcomes (Gottfredson, 1997a; 1997b). Efforts to create new or alternative types of broad intelligence, such as intuitively appealing models of learning styles, practical intelligence, or "street smarts", have not stood up to empirical tests in the way that CHC has (Cooper, 2015; Deary, 2001; Keith & Reynolds, 2010).

How Can They Be "Gifted" If They Don't Excel Academically?

Some teachers may expect gifted students to show high academic achievement, not just high ability or potential, before they can be identified. However, as we have discussed, factors such as early opportunity to learn, secondary learning challenges or exceptionalities, language spoken in the home, and ongoing home support all contribute to how a student performs academically. If, for example, a student is dyslexic, she may not achieve high performance or growth in reading. Lack of support or bearing greater responsibilities in the home can affect the achievement of some students more than others.

Motivation and academic engagement are also important to consider. Students with high ability may become disengaged from the curriculum and learning experiences used in the general education classroom. They may choose not to "play the game" of traditional learning, leading them to underperform, or they may have issues in their home life that affect motivation for school (Snyder & Wormington, 2020).

How Can a GT Identification from Elementary School Still Be Valid in High School?

This question reflects a common practice that should be challenged. Many schools assess students only once in elementary school and provide services throughout K-12 based on that early decision. Giftedness is really not a characteristic of the student, but a description of their current abilities compared to their peers which leads to different instructional needs. "One and done" identification is not appropriate because students can vary in their need for advanced instruction over time.

Gifted identification from elementary school should therefore not be automatically applied to decisions about coursework in high school or even middle school. More recent information should be used to allow students to be placed in the most appropriate services for their grade level (Lohman & Korb, 2006). A best practice is to regularly re-screen the full school population for students who need special academic challenge but did not show that need at the earlier assessment opportunity.

The age of the student is important, too, in deciding whether to rely on earlier identification results. For students testing in grades K, 1, or 2, it is especially important for scores to be recent. Young students grow and change in their readiness to learn relative to peers much more than older students. As mentioned earlier, environmental variables, such as opportunity to learn, the quality of the learning experiences, and family resources, affect students in the early grades more than later grades. Therefore, a student's score from K or 1 may likely be no longer valid to make programming decisions for grade 3.

Aren't Gifted Students Advanced in Every Academic Subject?

Should "profiles" of performance across different tests and subtests even be considered in the context of gifted education? Students with exceptional talents can have areas of relative weakness (e.g., verbal reasoning scores that are much lower than their quantitative reasoning scores). In fact, around 15% of high performing students on the CogAT have one area with markedly lower scores and only one-third are classified as "flat" profiles with no relative strengths or weaknesses across the verbal, quantitative, or nonverbal batteries (Lohman et al., 2008). This tendency is also true for talent search students of middle school age who are administered off-level tests. Thus, these profile differences should be important considerations in both selection and programming for potential.

The performance of students across multiple test batteries is called a *profile* and can reflect both the typical level of performance and the relatively stronger and weaker scores. Some tests offer interpretations of students' relative performance on different batteries or domains, which usually take measurement error into account (see earlier section on measurement error and confidence intervals). If educators seek to compare battery scores on their own, they should calculate confidence intervals for each score to take into account measurement error and ensure that differences in scores are large enough to be important. Comparisons should also be based on scaled scores intended to be compared across batteries, such as index scores or percentile ranks which are normally distributed. Number correct (raw scores) cannot be used to determine relative strengths.

Do We Need to Spend Time on Test Preparation?

Appropriate orientation to the format and style of the assessment should be used, especially for ability tests which are often more novel for students. This is especially important if better resourced or better connected parents are accessing test preparation for their students (Lu et al., 2020). Test familiarity follows the familiar learning curve, where students with no experience with testing benefit the most from a small amount of practice while students who have been extensively coached benefit little from more practice. Fortunately, this works in favor of students who lack test experience because providing a small amount of appropriate orientation to the test (e.g., pre-teaching or practice tests or brief lessons) will level the playing field for students with less experience without further benefit to students who are more familiar with testing.

Practical Examples of How Test Data Might Be Used in the Identification Process

Both ability and achievement tests are commonly used in gifted identification, but their specific value to the process can vary. Here, we highlight some common uses of achievement and/or ability scores in identifying students with current or potential for advanced academic performance.

Universal Screening for Services

One distinction made in the identification process is in the screening phase, where students are administered an assessment that determines if they may receive additional consideration, and the confirmation phase, where additional assessments or portfolios are conducted to make service decisions for individual students. Universal screeners are tests administered to an entire grade level for the purpose of recognizing students who need further consideration for advanced academic or gifted education services.

Whenever possible, universal screening is preferable to referral-led identification. Students from historically underrepresented groups or marginalized communities are much less likely to be referred by teachers who are not from that same group or community (Grissom & Redding, 2016). Parent referral or optional screening processes also create obstacles to identification that differentially exclude students whose parents are unable to provide transportation, do not understand the importance of screening, or are otherwise marginalized in their own relationship to schools. The best practice is to minimize the obstacles to being identified, which often means doing universal screening during the school day or using existing data (Yaluma & Tyner, 2018).

Naturally, universal screening is expensive if a test is used solely for this purpose, so it is common for schools to use existing scores, such as the annual, state-required achievement tests (Yaluma & Tyner, 2018). Ability tests, both uni- and multidimensional, are also commonly used in screening.

Because achievement tests are influenced by past learning opportunities, using them as the only universal screener will tend to perpetuate inequities. If this is the only screening test data available, programs should set lenient cutoffs for consideration (maybe considering students in the top half or third of achievement in the school; Peters et al., 2019), and teachers involved in screening for advanced academic services need to be proactive in looking for talent among students with relatively lower achievement who may have low motivation in the regular classroom or who have not had enrichment opportunities but could benefit from them.

Group-administered cognitive ability tests often are used for the initial screening phase of identification, including universal screening processes. The composite or total score from these tests often is used to determine eligibility for additional consideration. For unidimensional ability tests, this is the only score provided for this purpose. Multidimensional ability tests need to be combined in some way to make decisions about whether students require additional testing. These combinations can impact both the number and diversity of the identified students (Lakin, 2018a).

Ability tests also are used for the confirmation phase to provide a more precise assessment of academic need. Some districts use individually administered intelligence tests for this purpose while other districts use a multidimensional ability test or a portfolio of assessment data. Individually administered tests are the costliest to administer, because they require one-on-one testing time with a highly trained psychologist or psychometrist. Other districts use group-administered ability tests for confirmation when the initial screening is based on a shorter ability test or teacher referrals.

Quantitative Data as Part of a Portfolio or Body of Evidence (BOE) for Identification

Following screening, many districts use a more holistic and extensive evaluation process rather than a single confirmation test in order to align student needs to their service placement. Placement may be based on a matrix, which is a systematic approach to combining different types of information into a guided decision for services, or it could be based on a portfolio or case study approach where student services are determined based on professional judgment through a holistic review process (Callahan et al., 2018b; Moon, 2018). A holistic review is when multiple measures inform an expert decision, but no one piece of data or cut-score predominates the placement decisions (Mun et al., 2016; Worrell & Erwin, 2011).

This portfolio, or *body of evidence*, may be used for service determinations. These portfolios consist of normed-referenced cognitive data, academic data, and behavioral data alongside qualitative data. A review team, with at least one person trained or endorsed in gifted education and programming, reviews the BOE to determine gifted identification and programming needs. BOE are one approach to holistic placement decisions.

Matrix-Based Identification Decisions

When matrices are used to combine data, they should be carefully designed to avoid unintentionally creating cut scores or preventing students from receiving services in specific domains because they do not have universally high scores across all of the assessment tools (Lakin, 2018a; McBee et al., 2014; Moon, 2018). Callahan et al. (2018b) warned against the "smokescreen" (p. 89) that matrices create. First, each piece of data should inform the placement decision, rather than simply being a formality for consideration (e.g., a teacher nomination that simply determines if the student is tested). Second, only appropriate score scales should be used for comparisons. This may include percentile ranks (if normalized) or index scores that are intended to be contrasted across tests or batteries, but never raw scores. Third, the point values and relative weights applied to each source of data should be based on (a) the credibility of those scores, including their reliability and validity for the test interpretation and (b) the relevance of those characteristics to the service decision. Because of varying levels of performance across assessments, these scores should not be collapsed into a single point value (Moon, 2018). Finally, if any of the multiple measures makes the process too onerous for families or schools to maintain, it may decrease equity due to lack of participation in the process.

Cut-Scores

Cut-scores create an arbitrary division on a continuous scale. When used, cut-scores should be supplemented with expert judgment based on other evidence and consideration of measurement error. As we described earlier, measurement error is always present and students' range of plausible scores should be considered when using cut scores. A student scoring just a few points below a cut-off (on any assessment tool) would have a confidence interval that includes the cut-score and should have other opportunities to explore whether their true performance level demonstrates need. In other words, cut-scores should not be used firmly without additional consideration of students who just miss the score(s).

Identifying Students for a Developmental Talent Pool

Ideally, identification leads to a suite of potential services, not just one gifted program (Callahan et al., 2018b). Achievement and ability data can be used for Talent Pool identification, which is a level of services designed for students who are not yet ready for advanced academic courses but could become ready with additional support. Talent pool students may have achievement levels just outside of the qualifying range or have strong ability scores, especially when compared to peers with similar achievement levels. Different tiers of service may be defined for students showing this level of potential, where enrichment opportunities may allow students to accelerate their learning and qualify for services or advanced coursework in the future.

Acceleration

One essential use of achievement tests in gifted education is for acceleration decisions, either within an academic area or for whole-grade acceleration. The field of gifted education has provided strong evidence to counter misconceptions surrounding acceleration, which has been consistently shown to have net positive effects on students' academic achievement without serious social or emotional drawbacks (Assouline et al., 2017; Dare et al., 2019; Gross, 2006). Tools and guidelines have been developed to guide decisions about acceleration, such as the Iowa Acceleration Scale which provides for a systematic, data-based decision process for acceleration (Assouline et al., 2004).

Achievement tests should show a pattern of consistently meeting or exceeding grade level standards in multiple content areas to justify whole-grade acceleration. The authors of the Iowa Acceleration Scale (Assouline et al., 2004) suggest that above-level test results are the most informative data because they can show how students perform on content aligned to advanced grade levels. Students who perform at or above average on an above-level achievement test are more likely to be ready for whole-grade acceleration than students who only show exceptionalachievement on test content aligned to their current grade-level.

When making decisions about whole-grade acceleration, achievement in each academic domain needs to be evaluated to determine that the student is ready for advanced instruction in all areas. If a student demonstrates high achievement in just one or two academic areas, single-subject acceleration is an effective tool for serving those strengths. Decisions around single-subject acceleration should be made using battery or domain scores from the achievement test alongside measures of ability and interest that are also relevant to that domain.

Bringing Assessment Data into Service-Identification Alignment

In many school districts, gifted services have been siloed and separated from other services for learning differences. Because of that separation, and because of lack of funding, if a school or district offers gifted services, they often are "one size fits all," implying that all students who qualify need the same services. We would never provide identical services to students with different Individualized Education Plans (IEP), yet it is a common practice in gifted education (Gubbins et al., 2021).

The lack of coherence between the way gifted students are identified and the way they are served in gifted programs hampers the field's ability to show positive effects on students' academic growth (Adelson et al., 2012; Gubbins et al., 2021). Therefore, when evaluating the gifted identification processes or considering changes, it is important to consider the alignment of the measures to the programming options that are available (Lakin, 2018b).

If students are to be placed in an advanced math course, then it is vital to select students who have high math achievement. Using a screener that measures creativity or figural reasoning will not be sufficiently targeted. On the other hand, if the program offers compacted or accelerated instruction in math, *potential* for high achievement is just as important to assess as actual high achievement. In this case, the program might administer a quantitative reasoning assessment or engage students in performance assessments around mathematical inquiry. Their current math achievement might not even be used in placing them.

Alignment is also a vital consideration when multiple pathways are offered and especially when historically underrepresented groups are given alternative pathways to identification (Peters et al., 2014). Using different cut scores for different measures, even if aligned with equity goals, does not ensure equal readiness and academic need (NAGC, 2008). A student who has potential but not current high achievement in math may benefit from accelerated or more complex instruction on grade level in order to advance their learning and reach high academic achievement. However, a student with already high achievement may need above grade level content right away. If the goal is to place a more representative pool of students into an above-level course, then students identified by alternative pathways should be offered either different, more appropriate services, or bridge services to prepare them for the existing service (Wells, 2020).

Equity and Justice in Identification with Ability and Achievement Tests

Students arrive at the first day of kindergarten with great variation in their academic skills and readiness. These variations are largely due to family income levels and the opportunities socioeconomic resources can provide, including high-quality pre-kindergarten programs or differences in home life due to formal education, such as hearing more vocabulary words (García & Weiss, 2017). Unfortunately, racial and socioeconomic gaps in achievement tend to either stay the same or sometimes grow over time (Condron, 2009; Fryer & Levitt, 2004; Reardon, 2013).

Designing services not only for current high achievement, but also for potential, such as talent pool programs, can allow students with somewhat lower achievement scores to receive appropriate differentiation and begin to show higher level performance (Peters et al., 2014). Additional information is needed to identify students with *potential* for high achievement who have not received the right educational support or did not have the same enriched experiences as other students (Worrell & Erwin, 2011).

What about Using Different Tests?

A widespread misconception in the field of gifted identification is that there are types of reasoning tests that are better for identifying general academic talent for students from historically underrepresented or marginalized groups. No matter their background, academically successful students tend to have high ability in verbal, quantitative, and/or general reasoning ability (particularly compared to students with similar opportunity to learn), and they leverage these and other skills in their success in the classroom and in life alongside their personal characteristics such as interests and metacognitive skills (Borland, 2018; Callahan, 2009; Lakin & Lohman, 2011). The research has consistently shown that there is no assessment that can see through the effects of education and opportunity to measure *innate* abilities, and even tests intended to minimize cultural content are vulnerable to training and practice (Krautter et al., 2021; Schneider et al., 2020). As described earlier, all abilities are developed, and tests of any kind are a product of our culture and influenced by educational opportunity (Lohman, 2006). Tests themselves, of course, are a cultural invention.

What Are Developing Solutions to Equity in Identification?

A promising area of research is the use of "front-loading" or early enrichment programs (grades K, 1, or 2) where all students are given experiences with critical thinking and creativity skills before any kind of identification process is used. An example of this kind of program is the Young Scholars Program (Horn, 2015; Horn et al., 2021) and the curriculum discussed in Wells (2020). Such programs may improve equitable identification by either improving the screening scores of students from lower resourced families or historically marginalized groups or by creating additional opportunities for academic talent to be noticed and nurtured by the classroom teacher prior to identification (Briggs et al., 2008). More work is needed to explore these potential explanations.

Conclusion

Clearly, there is a great deal of nuance in understanding achievement and ability test scores and using them equitably and effectively in gifted identification. Assessment literacy is a skillset that more educators need to have (Johnsen, 2018), especially when making important data-based decisions including gifted identification.

Although many districts and states have responded to the call for multiple measures to be used in gifted identification (NAGC, 2019), the resulting data need to be thoughtfully considered and integrated into a coherent service model. Although the causes of group differences in achievement and ability scores are often beyond the control of educators (Peters, 2021), there are clear strategies to mitigate their effects in identification. This includes the use of gifted and talented universal screening, local, or subgroup norms, front-loading opportunities for learning the thinking skills being measured, and the use of professional judgment when cut-scores may artificially block students from receiving necessary services (Wells, 2020). Administrators engaged in gifted identification should heed the advice of many gifted experts to "Be a talent scout not a deficit detector" (North Carolina Department of Public Instruction, 2020; Renzulli, 1995).

References

Adelson, J. L., McCoach, D. B., & Gavin, M. K. (2012). Examining the effects of gifted programming in mathematics and reading using the ECLS-K. *Gifted Child Quarterly*, *56*(1), 25–39. http://doi.org/10.1177/0016986211431487

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing* (3rd ed.). Authors.

Assouline, S. G., Lupkowski-Shoplik, A., & Colangelo, N. (2017). Evidence overcomes excuses: Academic acceleration is an effective intervention for high ability students. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education* (pp. 173–186). Routledge.

Assouline, S. G., Colangelo, N., Ihrig, D., Forstadt, L., & Lipscomb, J. (2004) Iowa Acceleration Scale validation studies. In N. Colangelo, S. G. Assouline, & M. U. Gross (Eds.), *A nation deceived: How schools hold back America's brightest students.* (Vol. 2, pp. 167–172). Connie Belin & Jacqueline N. Blank International Center for Gifted Education and Talent Development. www.accelerationinstitute.org/nation_deceived/nd_v1.pdf

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge University Press.

Borland, J. H. (2018). Problematizing gifted education: Thinking radically about our beliefs and practice. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (2nd ed., pp. 71–82). Routledge.

Briggs, D. C., & Peck, F. A. (2015). Using learning progressions to design vertical scales that support coherent inferences about student growth. *Measurement: Interdisciplinary Research and Perspectives*, *13*(2), 75–99. https://doi.org/10.1080/15366367.2015.1042814

Briggs, C. J., Reis, S. M., & Sullivan, E. E. (2008). A national view of promising programs and practices for culturally, linguistically, and ethnically diverse gifted and talented students. *Gifted Child Quarterly*, *52*(2), 131–145. http://doi.org/10.1177/0016986208316037

Callahan, C. M. (2009). Myth 3: A family of identification myths: Your sample must be the same as the population. There is a "silver bullet" in identification. There must be "winners" and "losers" in identification and programming. *Gifted Child Quarterly*, *53*(4), 239–241. https://doi.org/10.1177/0016986209346826

Callahan, C. M., Renzulli, J. S., Delcourt, M. A. B., & Hertberg-Davis, H. L. (2018b). Identification of gifted and talented students. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (2nd ed., pp. 85–93). Routledge.

Carman, C. A., Walther, C. A., & Bartsch, R. A. (2018). Using the Cognitive Abilities Test (CogAT) 7 Nonverbal Battery to identify the gifted/talented: An investigation of demographic effects and norming plans. *Gifted Child Quarterly*, *62*(2), 193–209. https://doi.org/10.1177/0016986217752097

Condron, D. J. (2009). Social class, school and non-school environments, and black/white inequalities in children's learning. *American Sociological Review*, *74*(5), 685–708. https://doi.org/10.1177/000312240907400501

Conley, D. T. (2014). *The Common Core State Standards: Insight into their development and purpose.* Council of Chief State School Officers. https://csaa.wested.org/wp-content/uploads/2019/11/2014_CCSS_Insight_Into_Development.pdf

Conway, A. R., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J. P. (2021). Individual differences in attention and intelligence: A united cognitive/psychometric approach. *Journal of Intelligence*, *9*(3), 34. https://doi.org/10.3390/jintelligence9030034

Cooper, C. (2015). *Intelligence and human abilities: Structure, origins and applications*. Routledge.

Dare, L., Nowicki, E. A., & Smith, S. (2019). On deciding to accelerate: high-ability students identify key considerations. *Gifted Child Quarterly*, *63*(3), 159–171. https://doi.org/10.1177/0016986219828073

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford University Press.

Erwin, J. O., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment*, *30*(1), 74–87. http://doi.org/10.1177/0734282911428197

Foley-Nicpon, M., Assouline, S. G., & Colangelo, N. (2013). Twice-exceptional learners: Who needs to know what?. *Gifted Child Quarterly*, *57*(3), 169–180. http://doi.org/10.1177/0016986213490021

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447–464. https://scholar.harvard.edu/files/fryer/files/rest_vol86_2.pdf

García, E., & Weiss, E. (2017). *Inequalities at the school starting gate: Gaps, trends, and strategies to address them*. Economic Policy Institute. https://files.eric.ed.gov/fulltext/ED588751.pdf

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. https://doi.org/10.1016/S0160-2869(97)90014-3

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*(1), 1–25. http://doi.org/10.1177/2332858415622175

Gross, M. U. (2006). Exceptionally gifted children: Long-term outcomes of academic acceleration and nonacceleration. *Journal for the Education of the Gifted*, *29*(4), 404–429. https://doi.org/10.4219/jeg-2006-247

Gubbins, E. J., Siegle, D., Ottone-Cross, K., McCoach, D. B., Langley, S. D., Callahan, C. M., Bodersen, A.V., & Caughey, M. (2021). Identifying and serving gifted and talented students: Are identification and services connected?. *Gifted Child Quarterly*, *65*(2), 115–131. https://doi.org/10.1177/0016986220988308

Horn, C. V. (2015). Young scholars: A talent development model for finding and nurturing potential in underserved populations. *Gifted Child Today*, *38*(1), 19–31. https://doi.org/10.1177/1076217514556532

Horn, C.V., Little, C. A, Maloney, K., & McCullough, C. (2021). *Young scholars model: A comprehensive approach for developing talent and pursuing equity in gifted education*. Routledge. https://doi.org/10.4324/9781003239697

Johnsen, S. K. (Ed.). (2018). *Identifying gifted students: A practical guide* (3rd ed.) Prufrock Press.

Jonson, J. L., Trantham, P., & Usher-Tate, B. J. (2019). An evaluative framework for reviewing fairness standards and practices in educational tests. *Educational Measurement: Issues and Practice*, *38*(3), 6–19. https://doi.org/10.1111/emip.12259

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635–650. https://doi.org/10.1002/pits.20496

Koretz, D. (2009). *Measuring up: What educational testing really tells us.* Harvard University Press.

Krautter, K., Lehmann, J., Kleinort, E., Koch, M., Spinath, F. M., & Becker, N. (2021). Test preparation in figural matrices tests: Focus on the difficult rules. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.619440

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher*, *35*(7), 3–12. https://doi.org/10.3102/0013189X035007003

Lakin, J. M. (2018a). Making the cut in gifted selection: Selection decisions and their impact on program diversity. *Gifted Child Quarterly*, *62*(2), 210–219. http://doi.org/10.1177/0016986217752099

Lakin, J. M. (2018b). Service-driven identification practices: Why the services you offer must drive identification decisions. To appear in *Teaching for High Potential.* https://drive.google.com/open?id=17qoHlyrX_8_GZAow9EqFm4hweWromqQI

Lakin, J. M. (2020). Identification issues at a micro-level: Best practices regarding instrumentation and selection criteria. In Stambaugh, T., & Olszewski-Kubilius, P. (Eds.), *Unlocking potential: Identifying and serving gifted students from low income backgrounds* (pp. 79–108). Prufrock Press.

Lakin, J. M., & Lohman, D. F. (2011). The predictive accuracy of verbal, quantitative, and nonverbal reasoning tests: Consequences for talent identification and program diversity. *Journal for the Education of the Gifted*, *34*(4), 595–623. https://doi.org/10.1177/016235321103400404

Lohman, D. F. (1995). Developing academic talent: The roles of experience, mentoring, motivation, and volition. In N. Colangelo & S. G. Assouline (Eds.), *Talent development III: Proceedings from the Henry B. and Jocelyn Wallace national research symposium on talent development* (pp. 315–327). Gifted Psychology Press.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, *49*(2), 111–138. https://doi.org/10.1177/001698620504900203

Lohman, D. F. (2006). Beliefs about differences between ability and accomplishment: From folk theories to cognitive science. *Roeper Review*, *29*(1), 32–40. https://doi.org/10.1080/02783190609554382

Lohman, D. F., Gambrell, J., & Lakin, J. M. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science*, *50*(2), 269–282.

Lohman, D. F., & Korb, K. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted*, *29*(4), 451–484. https://doi.org/10.4219/jeg-2006-245

Lohman, D. F., & Lakin, J.M. (2017). *Cognitive Abilities Test (Form 8).* Riverside Insights.

Lu, Y., Weinberg, S. L., & McCormick, M. (2020). Test-taking for gifted and talented kindergarten: Underscoring the importance of outreach. *Gifted Child Quarterly*, *64*(4), 259–274. https://doi.org/10.3102%2F0013189X16633441

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345. https://doi.org/10.1111/j.1745-69-16.2006.00019.x

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science*, *27*(7), 1004–1018. https://doi.org/10.1177/0956797616644735

Matthews, M. S., & Peters, S. J. (2018). Methods to increase the identification rate of students from traditionally underrepresented populations for gifted services. In S. I. Pfeiffer, E. Shaunessy-Dedrick, & M. Foley-Nicpon (Eds.), *APA Handbook of Giftedness and Talent* (pp. 317–331). American Psychological Association. https://doi.org/10.1037/0000038-021

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69–89. https://doi:10.1177/0016986213513794

McGrew, K. S., & Schneider, W. J. (2018). *CHC theory revised: A visual-graphic summary of Schneider and McGrew's 2018 CHC update chapter*. MindHub publication #4 v1.2. www.iapsych.com/mindhubpub4.pdf

Moon, T. R. (2018). Uses and misuses of matrices in identifying gifted students: Considerations for better practice. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (2nd ed., pp. 116–124). Routledge.

Mun, R. U., Langley, S. D., Ware, S., Gubbins, E. J., Siegle, D., Callahan, C. M., McCoach, D. B., & Hamilton, R. (2016). *Effective practices for identifying and serving English language learners in gifted education: A systematic review of the literature* [report]. University of Connecticut. https://ncrge.uconn.edu/wp-content/uploads/sites/982/2016/01/NCRGE_EL_Lit-Review.pdf

National Association for Gifted Children & The Council of State Directors of Programs for the Gifted. (2015). *2014–2015 state of the states in gifted education: Policy and practice data* (report). www.nagc.org/resources-publications/gifted-state/ 2014-2015-state-states-gifted-education

National Association for Gifted Children Professional Standards Committee. (2019). *2019 Pre-K-Grade 12 Gifted Programming Standards*. www.nagc.org/ resources-publications/resources/national-standards-gifted-and-talented-education/pre-k-grade-12

National Council on Measurement in Education (2019, December 20). *NCME position statement: Misconceptions about group differences in average test scores*. www.ncme.org/blogs/megan-welsh1/2019/12/20/ncme-position-statement-misconceptions-about-group

North Carolina Department of Public Instruction (2020). *A call to action: Equity and excellence*. www.dpi.nc.gov/students-families/enhanced-opportunities/advanced-learning-and-gifted-education/call-action-equity-and-excellence

Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, *18*(11), 948–952. https://doi.org/10.1111/j.1467-9280.2007.02007.x

Peters, S. J. (2021). The challenges of achieving equity within public school gifted and talented programs [online first]. *Gifted Child Quarterly*. https://doi.org/10.1177/00169862211002535

Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, *60*(3), 159–171. https://doi.org/10.1177/F0016986216643165

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who gets served in gifted education? Demographic representation and a call for action. *Gifted Child Quarterly*, *63*(4), 273–287. https://doi.org/10.1177/0016986219833738

Peters, S. J., Makel, M. C., & Rambo-Hernandez, K. (2021). Local norms for gifted and talented student identification: Everything you need to know. *Gifted Child Today*, *44*(2), 93–104. https://doi.org/10.1177/1076217520985181

Peters, S., Matthews, M., McBee, M., & McCoach, D. B. (2014). *Beyond Gifted Education: Designing and Implementing Advanced Academic Programs*. Prufrock Press.

Rambo-Hernandez, K. E., & Warne, R. T. (2015). Measuring the outliers: An introduction to out-of-level testing with high-achieving students. *Teaching Exceptional Children*, *47*(4), 199–207. http://doi.org/10.1177/0040059915569359

Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership*, *70*(8), 10–16. https://eric.ed.gov/?id=EJ1029019

Renzulli, J. S. (1995). Teachers as talent scouts. *Educational Leadership*, *52*(4), 75–81. https://eric.ed.gov/?id=EJ496176

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018–2019 State of the States in Gifted Education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. www.nagc.org/2018-2019-state-states-gifted-education

Robins, J. H., & Jolly, J. L. (2018). Technical information regarding assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 73–116). Prufrock Press.

Ryser, G. (2018). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.) *Identifying gifted students: A practical guide* (3rd ed., pp. 33–58). Prufrock Press.

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

Siegle, D., Gubbins, E. J., O'Rourke, P., Langley, S. D., Mun, R. U., Luria, S. R., Little, C. A., McCoach, D. B., Knupp, T., Callahan, C. M., & Plucker, J. A. (2016). Barriers to underserved students' participation in gifted programs and possible solutions. *Journal for the Education of the Gifted*, *39*(2), 103–131. https://doi.org/10.1177/0162353216640930

Sireci, S. G. (2021). NCME Presidential Address 2020: Valuing educational measurement. *Educational Measurement: Issues and Practice*, *40*(1), 7–16. https://doi.org/10.1111/emip.12415

Snyder, K. E., & Wormington, S.V. (2020). Gifted underachievement and achievement motivation: The promise of breaking silos. *Gifted Child Quarterly*, *64*(2), 63–66. https://doi.org/10.1177/0016986220909179

Sternberg, R. J. (2011). The theory of successful intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 504–527). New York, NY: Cambridge University Press

Süß, H.-M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of measuring and understanding intelligence* (pp. 313–332). Sage.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, *97*(3), 484–492. http://doi.org/ 10.1037/0022-0663.97.3.484

Warne, R. T., & Larsen, R. (2021, January 5). School integration limits the ability of local norms to diversify gifted programs: A mathematical analysis with implications related to the achievement gap. https://doi.org/10.31234/osf.io/nemch

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Wells, A. (2020). *Achieving equity in gifted programming: Dismantling barriers and tapping potential*. Sourcebooks.

Worrell, F. C., & Erwin, J. O. (2011). Best practices in identifying students for gifted and talented education programs. *Journal of Applied School Psychology*, *27*(4), 319–340. http://doi.org/10.1080/15377903.2011.615817

Yaluma, C. B., & Tyner, A. (2018). Is *there a gifted gap? Gifted education in high-poverty schools*. (Research Report No. 1.31). https://fordhaminstitute.org/sites/default/files/publication/pdfs/0131-there-gifted-gap-gifted-education-high-poverty-schools.pdf

Zwick, R. (2019). Fairness in measurement and selection: Statistical, philosophical, and public perspectives. *Educational Measurement: Issues and Practice*, *38*(4), 34–41. https://doi.org/10.1111/emip.12299

CHAPTER

4

Qualitative Assessment Tools for Identification

Curriculum-Based Product/Performance Tasks

Catherine A. Little, Rebecca L. O'Brien, and Kelly L. Kearney

Recommendations around identification systems for gifted programs emphasize the use of multiple measures to support comprehensive assessment of students' strengths and needs. Among the recommended measures are tools that reflect students' regular academic work and evidence of potential from the classroom, not solely their performance on formal tests. Such classroom-based evidence might include products emerging from student work within the general education curriculum, or it might include evidence from specific performance assessment tasks conducted for the purpose of informing an identification process.

There are several important purposes for using curriculum-based products and performance tasks for identification for gifted services. Performance tasks and products tend to be representative of student work over time and in their general education learning context, thus providing a different type of evidence from what emerges in a testing environment (Worrell & Erwin, 2011). This connection to the general education learning context supports the predictive power of such tasks for how students may perform in a gifted program context (VanTassel-Baska, 2014), as well as demonstrating aspects of students' learning capacity (Calero et al., 2011). Multiple scholars have emphasized the potential of performance-based tasks and classroom products to be more inclusive and to provide more equitable opportunities for identification for students from traditionally underserved populations (Joseph & Ford, 2006; VanTassel-Baska, Johnson, & Avery, 2002; Zimmerman et al., 2020). A number of studies have demonstrated increased numbers of students from these groups referred for and identified for gifted programs based on performance on these types of tasks versus traditional measures (e.g., Sarouphim, 2001; VanTassel-Baska, Feng, & Evans, 2007; Zimmerman et al., 2020).

This chapter describes the use of curriculum-based products and performance tasks as part of a system of identification for gifted services. We explain several types of these identification tools, with examples of how they have been used in school settings to inform identification processes, with particular attention to the ways such tools may provide opportunities to support equity in identification and to increase access to gifted

DOI: 10.4324/9781003285991-6

services for students from populations that are typically underserved. We also describe key challenges and considerations in the use of these tools, including a focus on how they may fit within the total scope of an identification system and the benefits and drawbacks they may bring.

Curriculum-Based Products and Performance Tasks: Defining Types and Formats

There are several types of curriculum-based assessments that may be useful in identifying students' needs for gifted services and informing decisions around placement and planning. Curriculum-based products and performance tasks within an identification system can include both (a) assessment tools that are specifically administered as part of an identification process and (b) products emerging from general education classroom activities that are collected to inform an identification team's review of student data.

Such products from classroom activities can reflect aspects of students' content and conceptual knowledge, application of skills, and ability to communicate their understanding to a degree that standardized assessments often cannot (Karsenty, 2014; VanTassel-Baska, Johnson, & Avery, 2002). Often, products from classroom activities are less laden with the possibilities of test anxiety than standardized test results might be (Milligan, 2019). Further, the products from classroom activities may align more directly to the likely demands of a gifted program than other types of assessments, as well as directly reflecting student engagement with the grade-level curriculum. Thus, such products may provide the basis for predicting how students' strengths and needs may emerge within and be connected to the context of advanced learning programs.

Curriculum-Based Measures/Progress Monitoring Assessments

Many schools use curriculum-based assessments for regular progress monitoring of students within specific content areas. For example, schools frequently employ oral reading fluency (ORF) assessments to monitor student progress in reading, and several computer-adaptive assessments such as NWEA's MAP are used to monitor reading and math progress several times per year. These types of assessments supply evidence of growth and evidence of areas of need in major content areas, and in the case of the computer-adaptive tests in particular, they may provide aptitude information of student readiness for advanced-level work. These assessments have had somewhat less use in documenting a need for advanced-level services than for interventions for students who are not making expected progress; they frequently provide the basis for recommending that students receive additional academic support within schoolwide approaches such as Response to Intervention (RtI) or Multi-Tiered Systems of Support (MTSS) (McGowan et al., 2016). However, they hold the potential to demonstrate both high-level performance and patterns of growth that are relevant to identification decisions around gifted services (Johnsen & Sulak, 2013). McGowan et al. (2016) demonstrated that students identified for gifted services showed consistently higher scores on ORF assessments than students not identified, suggesting the potential utility of the tool as one data source for identification, although more research is needed to determine how these types of assessment may contribute to overall decision-making about identification.

Performance-Based Identification Tasks

Performance-based assessments, broadly defined, represent authentic tasks within a particular domain through which students are asked to apply a set of skills to a problem or question in context. They tend to represent a combination of skills and often draw on students' capacity for higher-level thinking. As an identification tool, when designed with sufficient space for students to respond at an advanced level, performance-based tasks may provide a strong basis for predicting what level of work students are capable of in response to advanced curriculum (VanTassel-Baska, 2014).

Some performance-based tasks for identification are *dynamic assessments*, meaning that they are designed to measure how students respond to specific instruction in a new area of focus or new skill; thus, the focus is measuring learning abilities "in action," or, related, measuring learning potential, not current achievement status (Calero et al., 2011; Cao et al., 2017). In many dynamic assessment contexts, teachers administer a pre-assessment to gather baseline data, then teach a related specific skill or strategy. Students engage with a task, using the newly learned skill, and those conducting the assessment measure the progress students have demonstrated from pre-assessment through the performance-based task (Calero et al., 2011). Thus, the focus of the assessment is on students' potential for learning, as opposed to a demonstration of learned skills. Because of this emphasis on learning potential, dynamic assessment has often been proposed as a method for supporting increased identification for students from underserved populations (Lidz & Macrine, 2001). For example, Chaffey and Bailey (2008) used a dynamic testing approach to identify high academic potential in children from Indigenous populations in Australia.

Similarly, other kinds of performance-based tasks with preteaching have also been used as a tool for identification, again often with the intent of addressing issues of equity and access for students from traditionally underserved populations. VanTassel-Baska, Johnson, and Avery (2002) spearheaded the development of a set of verbal and non-verbal performance-based tasks with a pre-teaching element for statewide administration as part of gifted program identification in South Carolina. Results demonstrated that higher proportions of students from underserved populations, specifically African-American students and those eligible for free and reduced lunch, were identified for gifted services using the tasks as compared to using only traditional measures of ability and achievement (VanTassel-Baska, Feng, & de Brux, 2007; VanTassel-Baska, Feng, & Evans, 2007; VanTassel-Baska, Johnson, & Avery, 2002).

Somewhat less formally, some schools may use observations of student engagement with "response lessons" or activities designed to draw out certain types of critical or creative thinking behaviors, particularly in the primary grades (Horn et al., 2021; Little et al., 2018; Milligan, 2019). Such a "talent spotting" approach, which usually involves collaboration between gifted resource teachers and classroom teachers to implement and observe high-level tasks and document students' responses, may be used as part of a screening system ahead of additional formal identification measures. This approach of using gifted curriculum to locate promising learners among underserved populations has been a common way to find students at the primary level for several decades (see VanTassel-Baska, Johnson, & Avery, 2002).

Work Products and Observations from Curriculum Tasks

Another common use of curriculum-based tasks in identification is the use of student products from classroom activities, collected after the fact, as evidence of achievement and potential. Teachers might be asked to submit samples of student work to be considered as part of a process of screening or identification. Such products are generally collected into a portfolio, assembled by teachers with or without input from students, and used as part of the case for decision-makers to consider around identifying students for gifted services. The difference between these products and those described above is primarily whether the purpose of the product was primarily to inform an identification process or whether it was part of regular classroom activities and is later applied to the identification process (Matthews, 2018), as well as the degree to which the administration and scoring processes are standardized for use for identification purposes.

Bridging this notion of student work products with the prior example of student engagement in specific response lessons to draw out behaviors, teachers might also specifically supply evidence of student engagement with advanced curriculum materials as part of a portfolio. When teachers implement advanced curriculum with all learners to invite demonstration of advanced learning behaviors, the evidence that emerges can form an important part of identification decisions. Moreover, again when such curriculum is implemented with all learners, there is increased opportunity for students from underserved groups to be recognized as having needs for gifted services (Robinson et al., 2018; Swanson et al., 2019).

This "talent spotting" approach with advanced curriculum can include observational assessment of student behaviors as well as collection of student products in response to the curriculum tasks. Observational components can provide a wealth of information about the student often unattainable through more standardized measures (Ready & Wright, 2011; Sudkamp et al., 2012, 2014). Observing students in everyday contexts as they engage with curriculum increases the probability that students will express the needs and abilities that necessitate identification (Chamberlain et al., 2007). Increased authenticity increases the validity of observational assessments of performance because it is more likely that the behaviors and skills to be identified as indicators of gifted potential are those being expressed and documented in everyday contexts. Critically, such approaches require that teachers have engaged in professional learning around *what* behaviors to look for and *how* to document them (Matthews & Peters, 2018; Swanson et al., 2019).

Role and Purpose of Curriculum-Based Product and Performance Tasks in Identification Systems

Curriculum-based tasks, collected from classroom activities, can form a valuable and informative component of an overall identification system for gifted programs. The National Association for Gifted Children (NAGC, 2019) maintains a set of *Pre-K–Grade 12 Gifted Education Programming Standards* to guide schools and districts with best practices related to gifted services. Within this resource, Standard 2 relates to assessment, and three of the assessment outcomes are specifically linked to identification systems and associated evidence-based practices. These practices include (a) a focus on using multiple components in an identification system, (b) examining the environment and the opportunities students have to demonstrate their potential, and (c) including data showing how students engage with above-level tasks, all of which connect closely with the types of curriculum-based data sources discussed here (see Appendix B for specific connections to the standard). In addition, the standards and practices signal the importance of professional learning around assessment tasks to ensure equitable access for students and attention to consistency in the use of various measures for identification purposes.

Curriculum-based tasks may be incorporated as part of an identification system at one or more key stages of the process, including during earlier screening steps or as part of a final identification decision component. Universal screening efforts usually include administration of some standardized assessment instrument with all learners, with metrics for identifying a screening group or talent pool. Engaging all learners in advanced curriculum or challenging response lesson opportunities and documenting their responses also represents an aspect of broadening access to the identification process and may be used in combination with more formalized assessments as part of a universal screening approach. Further, a talent pool of students screened through such universal approaches might then engage in further learning experiences and assessment activities as part of the process of identifying students for services, and within that talent pool component, curriculum-based tasks and products may be collected to inform the identification process.

Key Considerations in Using Curriculum-Based Product and Performance Tasks

The purpose of an assessment task, in general, is to elicit performance reflecting the person's current developmental state on a continuum of mastery within a content area or domain. The evaluation of performance carries high stakes, including decision-making regarding future instructional experiences, placements, or ratings of performance. Thus, assessment approaches and instruments must meet technical adequacy standards for validity, reliability (especially interrater reliability in the application of rubrics), and equity.

Validity refers to how well the assessment measures what it is intended to measure; in other words, the degree to which one can make a *valid inference* from the assessment result regarding the test-taker's current performance on a continuum of mastery. Reliability refers to the consistency and precision of the results. In the case of interrater reliability, it refers to the consistency across raters in application of rubrics and deriving scores for student performance. Equity refers to the avoidance of psychometric bias, or over- or underestimating scores by demographic group. It also refers to "layperson bias," or the perspectives of individuals involved that may influence psychological factors related to the assessment or to taking or giving the assessment (e.g., teachers' beliefs about the purpose of an assessment may affect how they implement that assessment in the classroom (Seden & Svaricek, 2018), or students' self-efficacy beliefs may influence performance on some types of assessment). The idea of fairness or equity in assessment also goes beyond technical bias to include a concern for the broader social context and influences (Camilli, 2006; Tierney, 2014). These questions of validity, reliability, and fairness influence the practical functionality and meaningfulness of an assessment's use for identification.

Essentially, curriculum-based products and performance tasks provide evidence of what students actually do during a learning activity. In a dynamic assessment context, this evidence reflects growth from a pre-assessment to a post-assessment performance with specific instructional intervention between them. With other types of performance-based tasks and products used for identification purposes, the evidence may reflect a pattern of student growth over the course of a particular unit or period of time, or a demonstration of student response to above-grade-level demands, or a sign of completed work that is exceptional for a student's age, background, experiences, or other factors. Within school settings, when performance-based tasks and products are considered as part of an identification system, educators must consider how they are applied and scored fairly and consistently, and how the information they provide supports valid inferences about students' need for and likely success in advanced learning programs.

There are several characteristics of curriculum-based tasks that make them useful for consideration as part of an identification system:

- *Linkage to relevant opportunities within gifted programming:* Depending on the nature of the gifted program and its specific services, curriculum-based tasks may provide more direct evidence of student progress in and readiness for relevant content and learning experiences than a standardized test. They are particularly relevant for showing readiness when they engage students in content and skills that are advanced for their grade level and/or the types of learning activities that might occur within the local gifted program.
- *Meaningfulness and engagement with the learning activity:* Because these types of tasks emerge from classroom-based activities, often directly in line with the general education curriculum, students may feel more connection to the work than they do to an isolated testing experience. Therefore, there may be more of a sense of meaning in student engagement with the task (Moon et al., 2005), which strengthens motivation and may provide a clearer picture of student potential. Further, because these measures are frequently directly tied to typical classroom activities, they may present less of a context for test anxiety for students than traditional assessments (Milligan, 2019).
- *Demonstration of the application of skills and intellectual behaviors:* Performance-based tasks tend to incorporate complex steps and variables, multiple aspects of problem finding and problem solving, and expectations for persistence and attention. Thus, the completion of these tasks provides evidence of student abilities beyond their achievement of specific content knowledge and skills (Moon, 2013; VanTassel-Baska, 2014).
- *Opportunities for demonstration of a wide range of abilities and behaviors:* Because of the extensive amount of data potentially available from the classroom learning environment, curriculum-based tasks provide some flexibility for those individuals who know students to select items that best demonstrate those students' skills and abilities. For example, when teachers (or students themselves) are selecting products for a portfolio as part of an identification case, there is some flexibility and opportunity to highlight the strongest evidence of a student's potential—though it is still critical to ensure that the selection of artifacts is comprehensive and representative of students' work in the areas relevant to program decisions.
- *Potential increase in equitable access to services:* As noted, a broad scope of abilities and behaviors may be captured in performance-based tasks and products, and they are frequently collected in authentic classroom contexts. Given the nature and context of such types of assessment, they are perceived to provide more opportunities for students from underserved populations to demonstrate their potential and be considered for gifted or advanced services—and have frequently resulted in identification of higher numbers of students from underserved groups than traditional assessments (Sarouphim, 2001; VanTassel-Baska, Feng, & Evans, 2007; Zimmerman et al., 2020).

There are also some potential drawbacks to these tasks as elements of an identification system:

- *Technical adequacy:* Questions of validity, reliability, and equity arise with curriculum-based tasks as with any other form of assessment. Curriculum-based tasks, particularly performance-based and product-based classroom tasks, generally have less technical data supporting their use than standardized measures (Moon, 2013), primarily because they tend to be locally developed and not to have had the same expectations for rigorous attention to reliability and validity evidence as part of their development (Matthews, 2018). There are exceptions to this generalization (see Ryser, 2018, and VanTassel-Baska, Johnson, & Avery [2002] for evidence of thoroughness in technical details in development of performance-based and other qualitative tasks), but in many cases the evidence provided within portfolio systems, for example, would not bear this history. Curriculum-based products and performance tasks also tend to have less consistency in administration and scoring than a standardized measure, and the flexibility that may be a strength for capturing a student's potential is a potential drawback in terms of consistent application of procedures. At the same time, curriculum-based tasks may support high levels of validity in terms of the inferences to be made about students' potential future performance in a gifted program, because of their connection to the likely learning experiences in such a setting (see Shaklee & Viechinicki [1995] and Johnsen & Ryser [1997] for validity studies).
- *Costs:* Although curriculum-based tasks are likely less expensive than published standardized assessments in terms of administration, they may add considerable costs with regard to the time and personnel resources required for review and scoring as part of the identification process (Moon, 2013), particularly in large school systems and if initial screening efforts cast a wide net to put together a talent pool of students.
- *Need for additional professional preparation:* Educators who are charged with administering assessments and implementing performance-based tasks will need the time, training, and resources to accomplish these efforts with fidelity. In addition, most classroom teachers have limited background in recognizing the needs of advanced learners and how high potential may look different in students from a wide range of backgrounds. For example, if classroom teachers are expected to collect portfolio items as a way of demonstrating student potential, they are likely to need professional learning support around how to make decisions about the types of artifacts that will accomplish that goal (Pfeiffer, 2015). Further, if teachers are expected to implement advanced curriculum to provide context for students to demonstrate advanced learning behaviors, they need support for professional learning around what constitutes advanced curriculum and how to use existing advanced materials with fidelity (Swanson et al., 2019).

Using Curriculum-Based Product and Performance Tasks in Practice: Key Questions

Given the potential benefits and drawbacks of curriculum-based performance tasks and products as contributors to an identification system, school and district leaders responsible for gifted program identification must consider a number of questions regarding how to use such measures. These questions, outlined below, include such issues as (a) where curriculum-based tasks fit within the larger scope of a screening and formal identification system for a specific gifted program; (b) how portfolios of evidence regarding specific students' work should be constructed; and (c) how review and scoring of these sources of evidence might be conducted.

Program Design and Outcomes and Implications for Assessment Tasks

Any identification system for gifted services should be integrally connected to the definition of giftedness that frames the program and the program goals. Thus, educators responsible for developing and implementing an identification system should carefully consider the program goals and intended outcomes for students and align identification measures with those goals and outcomes (Moon & Hock, 2020; Peters et al., 2020), guided by questions such as these:

- What is the definition of giftedness that guides the program, and what types of evidence are useful demonstrations of relevant abilities, potential, and achievements? What are the desired outcomes of the program?
- What are the specific indicators that the products or performance tasks will be used to demonstrate? Are these indicators related to developed skills, evidence of readiness or potential for advanced learning experiences, or some combination?
- Is the evidence that is provided through performance-based tasks different from other sources used, or is it corroborative?

If the gifted program is based on a definition that includes a strong emphasis on creativity, then some kind of measure of creativity should be part of the identification system, provided that creativity is a major emphasis of services as well. A gifted program that primarily focuses on advanced mathematics instruction, for instance, should not be using a performance-based writing assessment as a primary data source for identification. Identification measures, including performance-based tasks and products, should be selected based on the degree to which they help to distinguish students in need of and ready for the gifted services available.

In addition to these considerations linked to the content emphases reflected in gifted services, schools must also consider the degree to which their programs are intended to address a relatively broader focus on talent development based on emerging potential and not just on extending skills already evident at advanced levels; although those advanced level needs also require attention through identification and programming decisions. Thus, in addition to addressing demonstrated high ability and achievement, schools must examine the degree to which their programs are intended to promote and support the development of potential, particularly in students who may have had fewer past opportunities for such development. Such a focus on seeking and developing potential provides opportunities for schools to build programs that are more equitable and inclusive, and performance tasks and products may be useful components for identifying students for these programs, particularly because of the degree to which they provide evidence of learning *in progress*.

- To what degree will curriculum-based tasks be used as part of a *screening* stage for identification, and to what degree as part of formal identification for services?
- How will the performance tasks be integrated into the identification system, relative to other data sources?

Identification systems in a school or district often involve multiple stages in which student data are reviewed to inform decisions about placement for services. In general, there is some kind of screening phase—sometimes using a formal universal screening process, sometimes not—that establishes a "talent pool" of students to be considered for identification, and then further review occurs for students within that pool. In planning the use of performance-based tasks and products as part of identification, school teams must determine *when* such data sources are considered and *how* they are used in decision making. Some schools might choose to engage all students in an advanced curriculum experience and have teachers document students' work in that context as part of creating a talent pool (Horn et al., 2021; Swanson et al., 2019).

In other circumstances, performance-based tasks may be administered to the talent pool once students have been screened by other measures. For example, VanTassel-Baska, Johnson, and Avery (2002) developed performance-based assessments to be administered to students who had met an initial screening threshold on ability and achievement measures, and then these assessment results were used in combination with performance and other assessments for identification decisions. Because of the time-consuming nature of developing and reviewing portfolios of student work and/or administering performance-based assessments, these measures may be more practical following the establishment of an initial screening pool.

Schools must also determine how performance-based tasks or products inform the decision-making process in combination with other identification measures. For example, can students be determined eligible for gifted services based on ratings of a portfolio alone, or based on the combination of a portfolio with scores on other assessments? How are different elements weighed in the decision-making process? Matthews (2018) emphasized the value of qualitative measures in augmenting traditional assessments, particularly when these measures may help to inform identification decisions for students who are near a cutoff point on other assessments. Qualitative measures may be particularly useful in guiding identification committees to decisions that are more inclusive than otherwise for students who are near a cutoff point.

Providing students with multiple pathways to identification is likely to yield a more diverse group of students both demographically and in terms of areas of strength and need; yet requiring students to reach particular cutoffs on *all* measures used for identification will narrow the range (Lakin, 2018; McBee et al., 2014). At the same time, again, program goals should drive the identification process, and all identification measures should help to guide understanding of the student's readiness for and need for the program services. School/district identification teams must carefully consider how each data source adds validity to the placement decision and helps with informing the best placement for students in a defensible way (Moon, 2018); critically, a district team would examine multiple assessments with focus on patterns of strength that indicate where services are needed.

Constructing and Reviewing Portfolios of Student Work

A collection of classroom-based assessments and products in the form of a portfolio can be used by an evaluator to infer a student's knowledge, skill, and potential as a method of identification of a need for advanced academic services (Collins, 1992; Lam, 2014). To establish the utility and meaningfulness of the portfolio as an identification tool, an identification team must clearly define the purpose and clarify required elements and expectations (Collins, 1992; Riddle, 2017). Key considerations include the following questions:

- Will the portfolio be structured to demonstrate students' growth over time or sample achievement at particular points?
- Will the portfolio be expected to incorporate students' *typical* work or *best* work?

These questions are critical considerations for clarifying what a portfolio is supposed to represent about a student, because the answers would be vital for those assembling and reviewing the portfolio to know (Pfeiffer, 2015). "Best work" portfolios are more often used for identification purposes than "typical work" (Moon, 2013), but demonstrating growth through multiple products that increase in quality may be very informative for an identification review team.

- Who selects products for consideration?
- Are there specific artifacts that should be included in portfolios for all students being considered?
- Is there consistency in the way the artifacts are collected?
- How many artifacts are expected or allowed to be included in a portfolio? What guidance is provided to ensure they are representative of student performance?
- Will the products in a portfolio be accompanied by any explanation of or reflection on the work from the student or teacher for purposes of supporting interpretation?
- Who will be involved in reviewing portfolios or collections of student products, and how will those individuals be trained to ensure consistency in scoring?

Ideally, selection of works to be included in a portfolio should be decided collaboratively between teachers and students, and it is even more desirable if the teacher and/or student includes an explanation of the rationale for inclusion of those pieces (VanTassel-Baska & Hubbard, 2018). Portfolios can provide a wealth of knowledge about a student through a method that is more contextualized and authentic than traditional standardized assessments. Additionally, this authenticity and context specificity bolster the validity of the use of portfolios as a qualitative assessment tool (Lam, 2014). When the portfolio parameters of what is to be included as evidence become more explicit and scripted, that authenticity wanes. The reverse is also true, in that as the range and variety of what is included in a portfolio increases, the authenticity increases as well. In practice, it is more common that authenticity is high, and the parameters are less clear and ambiguous, which increases the challenges surrounding fairness and consistency in scoring. Thus, it is not surprising that some of the most repeated concerns about the use of portfolios focus on their validity, interrater reliability, and fairness (Koertz, 1998; Lam, 2016).

The apprehensions surrounding the validity and interrater reliability of portfolios as an assessment and identification tool are largely driven by whether (a) the portfolio provides an adequate representation about the student's performance such that accurate inferences can be drawn about their ability and potential for advanced work; (b) the evaluators are consistent in their ratings; (c) the level of challenge in the tasks and products designed by the teachers is judged to be advanced and comparable across classrooms; and (d) there is attention to the variability in the amount of adult assistance received by the student on the included products. These are understandable concerns and are often the major limits of successful large-scale use of portfolios (Koertz, 1998). In smaller scale use, such as at a selective or specialized school where teachers routinely collaborate on curriculum, portfolio assessment may be more manageable, in large part because there are a smaller number of trained evaluators who must work to achieve interrater reliability in their application of the review process.

A consistent concern is the method of scoring or evaluating the portfolios. With the inclusion of products from several content areas, it is likely that students' performances will vary greatly by subject area, particularly with the likelihood of asynchronous patterns of performance common in high potential learners. This issue underscores the importance of having clear scoring criteria and a guide to use in implementing the evaluation process. Ideally, each submission included in the portfolio should be evaluated separately, rather than arriving at a single score for the entire portfolio; such attention to individual artifacts provides a multi-faceted picture of students' strengths and needs, much like the broader use of multiple methods overall, and this attention supports the process of aligning students' strengths and needs to services (Simon & Forgette-Giroux, 2000). To further reduce scoring concerns, portfolio evaluators should be well trained on what constitutes exemplary performance to further reduce measurement and scoring errors. However, an understanding of these norms does little to account for the contextual influences and variability in students' backgrounds and experiences, both important considerations in the implementation of a fair and inclusive identification process.

Another scoring consideration is limiting the variability of a task's complexity and level of challenge that may occur across teachers. One solution is to require that all portfolios include products from the same tasks and activities. While this option might increase perceived validity and reliability, it reduces the inherent strengths of portfolio assessment—authenticity and equity. An alternative solution is requiring each submission within the portfolio to include a rating of the level of challenge of complexity of the task itself. By including ratings for the task in addition to scores for the submission itself, the realistic concern of task equity and comparability across teachers is reduced, and authenticity of the task is retained.

Traditional portfolio assessments include student reflections on the content of the portfolio and their reasoning for including particular submissions, a practice that contributes to the authenticity of portfolios and could address those individualized student experiences and contexts. Further, such reflections are also indicators of students' ability to evaluate their own work (Matthews, 2018), which may be valuable information to the overall identification process.

Another component that may also contribute to fair and inclusive identification is the addition of a reflection written by the teacher. The teacher's contribution would (a) provide perspective on important considerations in the student's background and context that might have contributed to the student's performance and (b) explain why and what aspects of the student's work should be considered exceptional for a student of similar background and experiences. Including this narrative is a strategy akin to using local norms in more quantitative methods of identification, to address concerns of inequity. It also provides the teacher an opportunity to communicate the level of support needed by the student and assistance required to produce the portfolio submissions, alleviating another of the main concerns surrounding portfolio assessment.

Sample Applications in School Settings

Once schools and districts have made decisions about the use of curriculum-based assessments for identification purposes, there are several ways such decisions can be put into action. A few specific examples of use of these assessments for identification systems are outlined here.

Talent Spotting through Advanced Curriculum

Engaging all learners in curriculum designed with advanced learners in mind is one way to provide opportunities for students from a wide range of backgrounds to demonstrate their capacity for responding to and engaging with advanced learning experience. Some schools and districts invite students to demonstrate their talents in response to advanced curriculum by expecting all teachers to implement at least one advanced learning unit per year, while also ensuring that classroom teachers have engaged in professional learning to prepare for implementing those units and recognizing emerging talent in their students (Horn et al., 2021). These approaches are often paired with identification stages that involve referrals from and/or input from classroom teachers as part of the identification process.

For example, Robinson et al. (2018) studied first grade classrooms in which teachers implemented challenging engineering curriculum with all learners as part of an effort to increase access to advanced services and provide a context for learners with high potential to show their learning capacity. Teachers who implemented these materials were asked after implementation to indicate which students they would be likely to refer for gifted services, and their responses showed that a higher number of students from underserved populations would be nominated than had been in prior years at the same schools. Similarly, Swanson et al. (2019) implemented a Talent Development Academy (TDA) at multiple Title I schools, through which teachers implemented rigorous curriculum with all learners, including a focus on specific teaching models and culturally relevant teaching. They found that the number of students nominated by TDA teachers substantially increased following engagement with the project.

Many advanced curriculum resources also include performance-based assessments that can be important indicators to consider in identification. For example, the curriculum materials in language arts and science from the Center for Gifted Education at William and Mary use performance-based tools that have been employed across multiple studies to assess achievement and curriculum effectiveness, and these tools could also be used as evidence of how students engaged with the learning and showed growth. The language arts units incorporate writing tasks with rubrics that have been tested across several studies for test-retest reliability and interrater reliability (Feng et al., 2004; VanTassel-Baska, Zuo, et al., 2002), and the science units use an experimental design assessment first introduced by Cain (1990) (Feng et al., 2004; VanTassel-Baska et al., 1998) that has also been used to demonstrate growth gains for learners. It is important to note, however, that these instruments have primarily been used for assessment of learning, not as identification instruments per se. They have been used in some districts, however, in combination with other measures, for identification purposes.

Response to Higher-Level Thinking Activities

Similar to the talent-spotting examples above, but generally following a less intensive time schedule than a full curriculum unit, educators might implement "response lessons" or activities that are intended to draw out critical and creative thinking behaviors that are potentially indicative of advanced potential (Horn et al., 2021; Little et al., 2018). Crucially, part of this process involves teacher collaboration, often a classroom teacher with a resource teacher, such that teachers share the responsibilities of implementing activities while observing for student behaviors. It also requires opportunities for professional learning and systematic approaches to documenting what teachers observe. Horn (2015) described the use of this approach as a critical component of the Young Scholars Model, which has resulted in increased numbers of students from underserved groups referred for and demonstrating success in advanced programs across several studies (Olszewski-Kubilius & Clarenbach, 2012; Little et al., 2018)

Another approach to encouraging students' responses to advanced activities is the use of model-eliciting activities (MEAs). MEAs are cooperative activities that focus on solving authentic, realistic mathematical problems while relying heavily on critical and creative thinking skills (Coxbill et al., 2013; Lesh & Yoon, 2004). Students develop generalizable mathematical models through "iterative cycles, expressing, testing, and revising their product, while verbalizing and justifying their own mathematical understanding" (Coxbill et al., 2013, p. 182). The benefit of these activities is that all students can access and perform the activity at their own readiness level, providing the necessary "room at the top" for students with advanced capabilities. Additionally, MEAs engage students in problems that are more sophisticated than often seen in standard curricula. MEAs have the potential to make students' mathematical thinking and reasoning more visible to the teacher, allowing for the observation and documentation of advanced (and often creative) performance within a low-risk, low-stakes context (Coxbill et al., 2013; Lesh et al., 2000).

Performance-Based Assessment Tasks

A few research and development projects have specifically involved the development, testing, and wider implementation of performance-based assessment tasks as part of identification systems for gifted programs, generally with a specific focus on increasing diversity in identification (e.g., Maker, 2005; Sarouphim, 2001; VanTassel-Baska, Johnson, & Avery, 2002; Zimmerman et al., 2020). VanTassel-Baska, Johnson, and Avery (2002) designed and tested a series of verbal, math, and nonverbal performance-based tasks for implementation as part of the statewide identification protocol in South Carolina. These tasks were designed to be open-ended, advanced for the grade level, and focused on higher-level thinking skills, and all included a pre-teaching component that would provide students with experience in the item type and teachers experience with higher level activities. They were based on the core components of the South Carolina curriculum standards in language arts and mathematics. The development team also wrote and tested rubrics for scoring the tasks. South Carolina professionals were trained in the scoring process and congregated in person to conduct the scoring process with monitors and checkers to enhance interrater reliability. Across several years of implementation, higher percentages (14% on average) of Black students and students from low-income backgrounds were identified through the performance tasks than through traditional assessments (VanTassel-Baska, Feng, & de Brux, 2007; VanTassel-Baska, Feng, & Evans, 2007).

The DISCOVER assessment, initially developed in the 1990s, consisted of five performance-based activities intended to assess aspects of students' problem-solving abilities across multiple domains (Maker, 2005). Observers use a checklist to score student responses to the tasks, and observers rotated such that each observer would score each child's response on only one task. Observers then met to discuss performance overall to inform identification decision-making. Across multiple studies, the DISCOVER assessment has demonstrated identification of higher percentages of students from underserved populations than were identified through traditional means (Sarouphim, 2001, 2004; Sarouphim & Maker, 2010). Some more recent work examining the use of performance-based assessments as part of identification for advanced work in STEM areas has similarly demonstrated that these assessments may yield identification of more students from underserved populations (Maker, 2020; Zimmerman et al., 2020).

Summary of Key Challenges and Recommendations

Curriculum-based performance tasks and products can provide identification teams in schools and districts with valuable information about how students engage with learning, their capacity for taking on advanced content and tasks requiring higher-level thinking, and how they show learning and growth over time, whether within the course of a dynamic task or over months represented in a portfolio. Notably, such tasks may broaden the population of learners who are considered for identification for gifted services, because of the range of skills and abilities they target as well as the increased authenticity of the conditions in which such assessments occur.

There are multiple advantages and disadvantages of using these types of approaches as part of an identification system, as outlined throughout this chapter. Educators considering performance-based tasks and products must weigh such questions as the balance of authentic representation of students' classroom work with consistency of measures across classrooms and schools. They must consider the training needs for the professionals who will administer and score performance-based tasks. They must also recognize and consider that while such assessments may promote more identification of students from underserved populations, they may also result in more identification of students who are *not* from underserved groups—thus, although students from underserved groups may increase in number identified, their proportions in programs may not increase simultaneously (VanTassel-Baska, Feng, & de Brux, 2007).

Despite these challenges, performance tasks and products are important options to consider because of the range of information they can provide, the potential for increasing access, and the opportunities for recognizing talent potential that may otherwise be missed. Professional learning, allocation of resources, and goal alignment are perhaps the most critical needs for ensuring the success of using performance-based tasks and student work products as part of an identification system. Educators need support to recognize advanced behaviors, implement learning activities that will yield relevant products and behaviors, and score assessments consistently and fairly. All aspects of identification should be considered in connection with the goals and focus of the gifted program for which students are being identified. Grounded in that critical understanding, performance tasks and classroom products can be closely linked to program goals and practices, potentially increasing the predictive power of these assessments for student success in the program. Moreover, because many of the types of tasks outlined are learning activities themselves, they provide not only evidence for assessors of student potential, but also preparation for students for their future program experiences—which also, ultimately, will contribute to student and program success.

Appendix B

Curriculum-Based Assessment Tools and NAGC Assessment Standard

Student outcomes within Standard 2: assessment	Selected evidence-based practices linked to outcomes	Connections to curriculum-based assessment tools as components of identification system
2.1 Identification. All students in Pre-K through grade 12 with gifts and talented have equal access to the identification process and proportionally represent each campus.	2.1.1 Educators develop environments and instructional activities that prepare and encourage students from diverse backgrounds to express characteristics and behaviors that are associated with giftedness.	Recognition that how students engage with instruction in the learning environment is important to the overall base of evidence for identification Importance of learning opportunities that allow students to express behaviors indicating advanced potential and talent
	2.1.3 Educators use universal screening and multiple indicators of potential and achievement at various grade levels from Pre-K through grade 12 to provide multiple entry points to services designed to meet demonstrated needs.	Acknowledgement of the importance of examining multiple indicators, including qualitative indicators Ensuring all students have opportunities to demonstrate talent potential through universal screening approaches, which may include evidence from curriculum-based assessments
2.2 Identification. Students with gifts and talents are identified for services that match their interests, strengths, and needs.	2.2.2 Educators select and use assessments that relate to services provided and identify abilities, interests, strengths, and needs based on current research.	Curriculum-based tasks may provide evidence of achievement and potential with direct connection to the focus of gifted services in a school or district.
	2.2.3 Educators use assessments that provide qualitative and quantitative information from a variety of sources.	Curriculum-based tasks provide more extensive qualitative information about student performance and engagement than more traditional identification measures such as standardized tests.
	2.2.4 Educators use assessments that provide information related to above-grade-level performance.	Curriculum-based tasks may be structured to incorporate above-level expectations for student performance. Such classroom tasks may be drawn from ungraded existing measures used in prior research or off-level tasks defined by state assessment benchmarks
	2.2.6 Educators have knowledge of student exceptionalities and collect assessment data while adjusting curriculum and instruction to learn about each student's developmental level and aptitude for learning (i.e., dynamic assessment).	Dynamic assessment approaches provide evidence of student capacity for learning and growth within the context of curriculum-based tasks.

Student outcomes within Standard 2: assessment	Selected evidence-based practices linked to outcomes	Connections to curriculum-based assessment tools as components of identification system
	2.2.7 Educators interpret multiple assessments in different domains and understand the uses and limitations of the assessments in identifying the interests, strengths, and needs of students with gifts and talents.	Recognition of the importance of ensuring professional readiness for the use of different types of assessment, including classroom tasks, as evidence of needs for gifted services
2.3 Identification. Students with identified gifts and talents represent diverse backgrounds.	2.3.1 Educators select and use equitable approaches and assessments that minimize bias for referring and identifying students with gifts and talents, attending to segments of the population that are frequently hidden or under-identified. Approaches and tools may include front-loading talent development activities, universal screening, using locally developed norms, assuring assessment tools are in the child's preferred language for communication or nonverbal formats, and building relationships with students to understand their unique challenges and needs.	Attention to the use of classroom tasks for "talent spotting" and the importance of ensuring access for all students to have opportunities to demonstrate strengths and potential

References

Cain, M. F. (1990). The diet cola test. *Science Scope*, *13*(4), 32–34. www.jstor.org/stable/45028712

Calero, M. D., Belen, G. M., & Robles, M. A. (2011). Learning potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. *Learning and Individual Differences*, *21*, 176–181. https://doi.org/10.1016/j.lindif.2010.11.025

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Praeger.

Cao, T. H., Jung, J. Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics*, *28*(3), 163–203. https://doi.org/10.1177/1932202X17714572

Chaffey, G. W., & Bailey, S. B. (2008). The use of dynamic testing to reveal high academic potential and underachievement in a culturally different population. *Gifted Education International*, *24*, 67–81. https://doi.org/10.1177/026142940802400109

Chamberlain, S. A., Buchanan, M., & Vercimak, D. (2007). Serving twice-exceptional preschoolers: Blending gifted education and early childhood special education practices in assessment and program planning. *Journal for the Education of the Gifted*, *30*(3), 372–394. https://doi.org/10.1177/016235320703000305

Collins, A. (1992). Portfolios for science education: Issues in purpose, structure, and authenticity. *Science Education*, *76*(4), 451–463. https://doi.org/10.1002/sce.3730760408

Coxbill, E., Chamberlain, S. A., & Weatherford, J. (2013). Using model-eliciting activities as a tool to identify and develop mathematically creative students. *Journal for the Education of the Gifted*, *36*(2), 176–197. https://doi.org/10.177/0162353213480433

Feng, A. X., VanTassel-Baska, J., Quek, C., Bai, W., & O'Neill, B. (2004). A longitudinal assessment of gifted students' learning using the integrated curriculum model (ICM): Impacts and perceptions of the William and Mary language arts and science curriculum. *Roeper Review*, *27*(2), 78–83. https://doi.org/10.1080/02783190509554294

Horn, C. V. (2015). Young Scholars: A talent development model for finding and nurturing potential in underserved populations. *Gifted Child Today*, *38*(1), 19–31. https://doi.org/10.1177/1076217514556532

Horn, C.V., Little, C. L., Maloney, K. N., & McCullough, C. A. (2021). *Young Scholars Model: A comprehensive approach for developing talent and pursuing equity in gifted education.* Prufrock Press.

Johnsen, S. K., & Ryser, G. R. (1997). The validity of portfolios in predicting performance in a gifted program. *Journal for Education of the Gifted*, *20*(3), 253–267. https://doi.org/10.1177/016235329702000304

Johnsen, S. K., & Sulak, T. N. (2013). Screening, assessment, and progress monitoring. In M. R. Coleman & S. K. Johnsen (Eds.), *Implementing RtI with gifted students: Service models, trends, and issues* (pp. 23–46). Prufrock Press.

Joseph, L. M., & Ford, D.Y. (2006). Nondiscriminatory assessment: Considerations for gifted education. *Gifted Child Quarterly*, *50*(1), 42–51. https://doi.org/10.1177/001698620605000105

Karsenty, R. (2014). Mathematical ability. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 372–375). Springer. https://doi-org.proxy-remote.galib.uga.edu/10.1007/978-94-007-4978-8

Koertz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy, & Practice*, *5*(3). https://doi.org/10.1080/0969595980050302

Lakin, J. M. (2018).Making the cut in gifted selection: Score combination rules and their impact on program diversity. *Gifted Child Quarterly*, *62*(2), 210–219. https://doi.org/10.1177/0016986217752099

Lam, R. (2014). Promoting self-regulated learning through portfolio assessment: Testimony and recommendations. *Assessment & Evaluation in Higher Education*, *39*(6), 699–714. https://doi.org/10.1080/02602938.2013.862211

Lam, R. (2016). Taking stock of portfolio assessment scholarship: From research to practice. *Assessing Writing*, *31*, 84–97. http://dx.doi.org/10.1016/j.asw.2016.08.003

Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T. (2000). Principles for developing thought-revealing activities for students and teachers. In A. Kelly & R. Lesh (Eds.), *Research design in mathematics and science education* (pp. 591–646). Lawrence Erlbaum.

Lesh, R., & Yoon, C. (2004). Evolving communities of mind: In which development involves several interacting simultaneously developing strands. *Mathematical Thinking and Learning*, *6*, 205–226. https://doi.org/10.1207/s15327833mtl0602_7

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners. *School Psychology International*, *22*(1), 74–96. https://doi.org/10.1177/01430343010221006

Little, C. A., Adelson, J. L., Kearney, K. L., Cash, K., & O'Brien, R. (2018). Early opportunities to strengthen academic readiness: Effects of summer learning on mathematics achievement. *Gifted Child Quarterly*, *62*(1), 83–95. https://doi.org/10.1177/0016986217738052

Maker, C. J. (2005). *The DISCOVER project: Improving assessment and curriculum for diverse gifted learners.* Technical report RM05206. National Research Center on the Gifted and Talented.

Maker, C. J. (2020). Culturally responsive assessments of spatial analytical skills and abilities: Development, field testing, and implementation. *Journal of Advanced Academics*, *31*(3), 234–253. https://doi.org/10.1177/1932202X20910697

Matthews, M. S. (2018). Utilizing non-test assessment in identifying gifted and talented learners. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (2nd ed., pp. 135–145). Routledge.

Matthews, M. S., & Peters, S. J. (2018). Methods to increase the identification rate of students from traditionally underrepresented populations for gifted services. In S. I. Pfeiffer, E. Shaunessy-Dedrick, & M. Foley-Nicpon (Eds.), *APA handbook of giftedness and talent* (pp. 317–331). American Psychological Association. https://doi.org/10.1037/0000038-021

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69–89. https://doi.org/10.1177/0016986213513794

McGowan, M. R., Runge, T. J., & Pedersen, J. A. (2016). Using curriculum-based measures for identifying gifted learners. *Roeper Review*, *38*(2), 93–106. https://doi.org/10.1080/02783193.2016.1150376

Milligan, J. L. (2019). *Assessment of giftedness: A concise and practical guide* (3rd ed.). YBK.

Moon, T. R. (2013). Not just a test: Utilizing non-test assessments in identifying gifted and talented students. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (pp. 148–151). Routledge.

Moon, T. R. (2018). Uses and misuses of matrices in identifying gifted students: Considerations for better practice. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (2nd ed., pp. 116–124). Routledge.

Moon, T. R., Brighton, C. M., Callahan, C. M., & Robinson, A. (2005). Development of authentic assessments for the middle school classroom. *The Journal of Secondary Gifted Education*, *16*(2–3), 119–133. https://doi.org/10.4219/jsge-2005-477

Moon, T. R., & Hock, M. (2020). Alternative and nonverbal assessments. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: A survey of current research on giftedness and talent development* (3rd ed., pp. 321–333). Prufrock Academic Press.

National Association for Gifted Children. (2019). *2019 Pre-k-grade 12 gifted programming standards*. Author.

Olszewski-Kubilius, P., & Clarenbach, J. (2012). *Unlocking emergent talent: Supporting high achievement of low-income, high-ability students*. National Association for Gifted Children.

Peters, S., Ottwein, J., Lee, L. E., & Matthews, M. S. (2020). Identification. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: A survey of current research on giftedness and talent development* (3rd ed., pp. 261–272). Prufrock Academic Press.

Pfeiffer, S. I. (2015). *Essentials of gifted assessment*. Wiley.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335–360. https://doi.org/10.3102/0002831210374874

Riddle, S. (2017). Ecological congruence and the identification of learning disabilities. *Child Youth Care Forum*, *46*, 161–174. https://doi.org/10.1007/s10566-016-9376-8

Robinson, A., Adelson, J. L., Kidd, K. A., & Cunningham, C. M. (2018). A talent for tinkering: Developing talents in children from low-income households through engineering curriculum. *Gifted Child Quarterly*, *62*(1), 130–144. https://doi.org/10.1177/0016986217738049

Ryser, G. R. (2018). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 33–57). Prufrock Press.

Sarouphim, K. M. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly*, *45*(2), 130–138. https://doi.org/10.1177/001698620104500206

Sarouphim, K. M. (2004). DISCOVER in middle school: Identifying gifted minority students. *The Journal of Secondary Gifted Education*, *15*(2), 61–69. https://doi.org/10.4219/jsge-2004-446

Sarouphim, K. M., & Maker, C. J. (2010). Ethnic and gender differences in identifying gifted students: A multi-cultural analysis. *International Education*, *39*(2), 42–56. https://eric.ed.gov/?id=EJ930085

Seden, K., & Svaricek, R. (2018). Teacher subjectivity regarding assessment: Exploring English as a foreign language teachers' conceptions of assessment theories that influence student learning. *Center for Educational Policy Studies Journal*, *8*(3), 120–139. https://doi.org/10.26529/cepsj.500

Shaklee, B., & Viechnicki, K. (1995). A qualitative approach to portfolios: The Early Assessment for Exceptional Potential Design. *Journal for the Education of the Gifted*, *18*(2), 156–170. https://doi.org/10.1177/016235329501800204

Simon, M., & Forgette-Giroux, R. (2000). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research, and Evaluation,* 7(1), Article 18. https://doi.org/10.7275/bh4d-me80

Südkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Südkamp, A., Kaiser, J., & Moeller, J. (2014). Teachers' judgments of students' academic achievement. In S. Krolak-Schwerdt, S. Glock, & M. Boehmer (Eds.), *Teachers' professional development: Assessment, training, and learning* (pp. 5–25). Sense.

Swanson, J. D., Russell, L. W., & Anderson, L. (2019). A model for growing teacher talent scouts: decreasing underrepresentation of gifted students. In S. R. Smith (Ed.), *Handbook of giftedness and talent development in the Asia-Pacific* (pp. 1–20). Springer.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, *43*, 55–69. http://dx.doi.org/10.1016/j.stueduc.2013.12.003

VanTassel-Baska, J. (2014). Performance-based assessment: The road to authentic learning for the gifted. *Gifted Child Today*, *37*(1), 41–47. https://doi.org/10.1177/1076217513509618

VanTassel-Baska, J., Bass, G., Ries, R., Poland, D., & Avery, L. D. (1998). A national study of science curriculum effectiveness with high ability students. *Gifted Child Quarterly*, *42*(4), 200–211. https://doi.org/10.1177/001698629804200404

VanTassel-Baska, J., Feng, A. X., & de Brux, E. (2007). A study of identification and achievement profiles of performance task-identified gifted students over 6 years. *Journal for the Education of the Gifted*, *31*(1), 7–34. https://doi.org/10.4219/jeg-2007–517

VanTassel-Baska, J., Feng, A. X., & Evans, B. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly*, *51*(3), 218–231. https://doi.org/10.1177/0016986207302717

VanTassel-Baska, J., & Hubbard, G. F. (2018). Learning assessments for gifted learners. In J. L. Roberts, T. F. Inman, & J. H. Robins (Eds.), *Introduction to Gifted Education* (pp. 165–182). Prufrock Press.

VanTassel-Baska, J., Johnson, D., & Avery, L. D. (2002). Using performance tasks in the identification of economically disadvantaged and minority gifted learners: Findings from Project STAR. *Gifted Child Quarterly*, *46*(2), 110–123. https://doi.org/10.1177/001698620204600204

VanTassel-Baska, J., Zuo, L., Avery, L. D., & Little, C. A. (2002). A curriculum study of gifted student learning in the language arts. *Gifted Child Quarterly*, *46*(1), 30–44. https://doi.org/10.1177/001698620204600104

Worrell, F. C., & Erwin, J. O. (2011). Best practices in identifying students for gifted and talented education programs. *Journal of Applied School Psychology*, *27*(4), 319–340. https://doi.org/10.1080/15377903.2011.615817

Zimmerman, R. H., Maker, C. J., & Alfaiz, F. (2020). Culturally responsive assessment of life science skills and abilities: Development, field testing, implementation, and results. *Journal of Advanced Academics*, *31*(3), 329–366. https://doi.org/10.1177/1932202X20923981

CHAPTER

5

Use of Teacher Rating Scales to Augment Identification

Gail R. Ryser

Rating scales are often used in surveys to make inferences from a sample to a population. They are also used to make decisions about accountability and for program services. A major advantage of using rating scales to determine services is that they allow educators to combine many observations of students' behavior in an efficient way (Jarosewich et al., 2002). In this chapter, I will examine the use of rating scales to assist in the placement of students in gifted and talented programs. I will also discuss the characteristics of psychometrically sound rating scales and how to best use them so the principles of fairness in testing are met.

Definition of Rating Scales

Rating scales consist of statements or questions that have a close-ended response format with predetermined anchors to which individuals (e.g., educators, parents) provide responses (Katz & Rudolph, 2018). Anchors are developed based on the information that is intended to be inferred from the rating scale's scores. Anchors often use a Likert scale (Likert, 1932), which consists of five ordered responses from "strongly agree/approve to strongly disagree/disapprove." Sometimes the Likert scale is modified by users to include other descriptors for the anchors such as "never" to "always" or to include a larger range of response options from five to as many as 11. (See Figure 5.1.) Rating scale items should provide a consistent unit of measurement across all statements or questions.

The National Association for Gifted Children's (NAGC) *Pre-K–Grade 12 Gifted Programming Standards* (NAGC, 2019) presents several standards related to assessment, noting that both qualitative and quantitative assessments should be used to identify students for gifted programming. Qualitative assessments use words to describe an individual's strengths, while quantitative assessments use numbers to describe an individual's strengths. In addition, there are differences between the two types of assessments in the degree to which the assessment is dynamic or static and the degree to which the assessment reflects performance in the real world. Qualitative measures are more dynamic and reflect performance in the real world while quantitative measures

DOI: 10.4324/9781003285991-7

Behavior	Rating				
The student	Never	Rarely	Some	Somewhat More	Much More
1. Uses creativity in solving problems.					
2. Uses above level vocabulary.					

FIGURE 5.1 Example of anchors

are more static and typically consist of selected response options (Ryser, 2018a). Rating scales overall are considered qualitative assessments because they are based on observations of behaviors.

The *Pre-K–Grade 12 Gifted Programming Standards* (NAGC, 2019) also recommend that multiple indicators of potential and achievement be used to provide multiple entry points for services to meet gifted students' needs. Using multiple criteria such as teachers' ratings and test score information is one of the most common ways of identifying students as gifted because using multiple criteria is thought to improve the selection process (Acar et al., 2016; Klein & Fodor, 2019; Westberg, 2012). Rating scales (either completed by teachers and/or parents) are often included in the identification of gifted students because they provide qualitative information on the behaviors of students, which can be observed in their everyday environment (Klein & Fodor, 2019).

Characteristics of Rating Scales in the Identification of Students as Gifted

When using rating scales for referral and identification of students as gifted, ratings should be based on observations of behavior, not vague descriptions of characteristics. Behaviors should consist of a broad sampling of the most important characteristics of students who have talent, and the behaviors described must be observable in the setting in which the rating scale is used. When the constructs being measured are observable in multiple settings (e.g., school, home, after school activities), educators should gather information from multiple sources (e.g., parents, teachers, peers) that allows for a more complete understanding of students' strengths. Multiple raters can be from different settings (e.g., school and home) or from the same setting (e.g., two or more educators rate the student using the same scale and come to consensus about the ratings). Finally, when teachers are trained in the characteristics of gifted and talented students and the use of the rating scale the district plans to use to screen gifted students, teachers are more likely to nominate a more diverse pool of students. Renzulli et al. (2010) present a method for training teachers in the use of rating scales. He and his co-authors suggested that examiners provide teachers with the rating scale and ask them to categorize the items according to the key concepts in each of the constructs being measured. Teachers then work in small groups to list behaviors that are indicators of the items. Training assists teachers in understanding what behaviors are important when rating students.

Using Rating Scales to Identify Students for Gifted Programs

Most school districts use multiple measures when identifying students as gifted, and these measures are often administered along with a teacher or parent referral. Callahan et al. (2013) conducted a study of the status of elementary gifted programs in the United States. Of those districts who responded, the researchers reported that 86.5% used teachers' referrals and 80.5% used parents' referrals as the first step in the process of identifying gifted students. These referrals were often non-standardized which implies that teachers and/or parents were asked to refer students for gifted programs but were not provided with well-developed criteria. Using a standardized and well-designed rating scale can result in more valuable and consistent referrals from teachers and/or parents. Rothenbusch et al. (2018) found that teachers and parents are both important sources when gathering information about talent and that the accuracy of their ratings do not differ statistically between these two groups. In a meta-analysis, Acar et al. (2016) found that the consistency between nonperformance measures (e.g., parent referrals, teacher rating scales) and performance measures (e.g., cognitive ability tests, achievement tests) was higher for teacher rating scales versus other nonperformance measures. They further recommended that because the correlations between teacher ratings and performance measures were in the medium range, teacher ratings should be used to support identification of gifted students in conjunction with performance measures. Many researchers (e.g., Gentry & Mann, 2008; Johnsen, 2018; Peters & Gentry, 2012; Renzulli et al., 2010; Worrell & Erwin, 2011) agree that teachers' and/or parents' ratings of talent are particularly useful when rating scales use observable behavioral indicators in various areas over time and/or when those who use them are trained in their use. Rating scales with good psychometric properties are also effective tools for gathering information about constructs associated with giftedness that are more subjective and less often assessed, such as creativity and leadership.

The Issue of Underrepresentation

The field of gifted education has long recognized that low-income, culturally/linguistically diverse, and twice-exceptional students are underrepresented in gifted programs (e.g., Card & Guiliano, 2016; Ford, 1998; Peters et al., 2019; Ricciardi et al., 2020). Some researchers have made progress in how teacher ratings should be used to increase the proportion of underrepresented students in gifted programs. For example, in a study by Peters and Gentry (2012), low-income students were identified for gifted programs in proportion to their population in a school district when teacher ratings used district group norms and combined these ratings with achievement scores. Group norms are calculated using a specific subgroup, in this study, low-income students within the district. Hunsaker et al. (1997) examined if teacher nominations of culturally diverse or low-income students were more useful if they were based on an instrument designed around traits, aptitudes, and behaviors of giftedness. The rating instrument used in the study was the Traits Attributes Behaviors (TABS) Summary Form (Fraiser et al., 1995), which teachers used to refer culturally diverse or low-income students for gifted program placement. The study examined the relationship between teacher ratings and performance in gifted programs as measured by the Scale for Rating Students' Participation in the Local Gifted Education Program (Renzulli & Westberg, 1991). The authors found that teacher ratings were predictive of successful performance in some aspects of the gifted program. Specifically, they found that the ratings were predictive of creativity and social skills (i.e., group skills, language abilities, and enthusiasm). While the ratings were not predictive of more academic variables, the authors found promise that with training, teachers were able to identify culturally diverse and low-income students who were successful in specific non-intellective aspects of gifted programs.

Minimizing Bias and Increasing Fairness in Testing

To improve representation in gifted education programs, educators need to consider three persistent issues: low educator expectations, exclusive definitions, and lack of test fairness (Ryser, 2018b). Educators may exhibit deficit thinking that results from a lack of understanding of how characteristics may vary across racial and cultural groups (Moore et al., 2005). Professional learning is needed to increase educators' knowledge of the characteristics of culturally and linguistically diverse students, particularly those from low income backgrounds. Moreover, a definition that focuses on students who perform two or more standard deviations above the mean (i.e., 130) and primarily on quantitative measures, such as intelligence and achievement tests, also may exclude students from diverse backgrounds. Using multiple assessments and more inclusive criteria will often solve the challenges with narrow definitions. According to the testing standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), fairness in testing minimizes barriers so those who use test results can make valid score interpretations for the widest range of individuals as possible. There are four important principles or clusters of fairness, which all focus on the validity of score interpretation for a test's intended uses. These four clusters are:

- technical aspects of the test (i.e., test design, administration, and scoring)

A rating scale that is used by parents should be available in other languages or should allow users to translate items to other languages as needed. The constructs being measured by the rating scale should be clearly delineated and consist of behaviors that are observable in the school and/or home setting. Students must have the opportunity to exhibit the behaviors that are rated.

- validity of test score interpretations

The test or rating scale should not differentially favor one subgroup over another (e.g., low-income students). Users of rating scales will want to ensure that the technical manual includes a study of item bias and reports reliabilities by subgroups. In addition, a study of the diagnostic accuracy of gifted students for the rating scale provides additional evidence that a rating scale meets the standards. Diagnostic accuracy is a study in which the sensitivity(i.e., how well the measure identifies those who are likely to be gifted) and specificity (i.e., how well the measure identifies those who are not likely to be gifted) are examined at several standard score cut-offs.

- accommodations to remove construct-irrelevant barriers

It is critical that examiners remove barriers that would interfere with examinees' ability to demonstrate the behaviors on the constructs measured by teacher and/or parent rating scales. In large part this is achieved by making a test as accessible as possible to examinees through universal design and making adaptations as needed for learners with special needs.

- safeguards to inappropriate uses of the test scores

Examiners should use rating scales' scores for the purposes for which they are intended, match the program's services, and consider alternative explanations for test performance when warranted. For example, when using a rating scale measuring several constructs, educators should consider that students might be twice exceptional if they score unevenly, in the range of gifted on one subscale but not another.

Appendix C provides a complete description of current rating scales used to identify gifted students. These instruments all meet the criteria for being psychometrically sound and appropriate for most districts to use as part of their identification system if they meet the content criteria set by the district for program participation. The six scales are:

- The Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003)
- The Gifted and Talented Evaluation Scales-Second Edition (GATES-2; Gilliam & Jerman, 2015)
- The HOPE Teacher Rating Scale (Gentry et al., 2015) (designed to focus on high-potential students from low-income families)
- The Scales for Identifying Gifted Students-Second Edition (SIGS-2; Ryser et al., 2021)
- The Scales for Ratings of the Behavioral Characteristics of Superior Students-Third Edition (SRBCSS-III; Renzulli et al., 2010)
- The Universal Talented and Gifted Screener (UTAGS; McCallum & Bracken, 2018)

Interpretation of Data

While combining teacher and/or parent ratings with other information is imperative, more importantly educators should consider the rich qualitative information obtained when using a rating scale. All too often a selection committee is only provided the standard score for a norm-referenced rating scale. This in effect, makes rating scales quantitative because the information concerning the types of behaviors and strengths students exhibit is lost when one only has access to a single standard score (Ryser, 2018a). For a rating scale (or any type of nonperformance measure) to be truly qualitative, a selection committee must have access to the specific types and examples of behaviors that are rated highly. For example, some rating scales ask raters to provide examples of how students exhibit strengths. These examples should be made available to the selection committee.

Johnsen (2018) provides five important guidelines for combining data from various assessments including rating scales to make decisions about identification of students as gifted. These guidelines are:

- each assessment should be given equal value in the selection process.
- all assessments scores should be comparable (i.e., raw scores cannot be compared to standard scores).
- measurement error should be considered.
- best performance should be reported as it is indicative of the student's potential.
- anecdotal information or clinical observations should be reported.

Name: Tamara Garcia

DOB: 4/22/2012. Age: 7

School: Lincoln Elementary

Assessments Used (Criterion = 125 standard score based on a mean of 100 and a standard deviation of 15)

Name of Assessment	**Raw Score**	**Standard Score**	**68% CI**	**Criterion Met**		
				Yes	**w/SEM**	**No**
SIGS-2						
GIA	32	121	118–125		X	
Mathematics	38	136	133–139	X		
SAGES-3						
Mathematics/Science	21	133	128–138	X		
Nonverbal Reasoning	20	124	120–128		X	

FIGURE 5.2 A sample student profile sheet

The last guideline is particularly important when one uses teacher and/or parent rating scales that are based on behavioral observations as the qualitative information often provides a compelling case for the student's success in the gifted and talented program.

Figure 5.2 contains an example of combining information from assessments, including rating scales, to make placement decisions in gifted programs. In my example I use the Scales for Identifying Gifted Students-Second Edition School Rating Scale (SIGS-2 SRS; Ryser, et al., 2021) and the Screening Assessment for Gifted Elementary and Middle School Students-Third Edition (SAGES-3, Johnsen & Corn, 2019).

Qualitative Information: Tamara excels in mathematics. She understands new mathematical concepts quickly and enjoys playing with numbers. She can generalize from one mathematical problem to another. She has good spatial ability and loves to make models. Her favorite subject is mathematics, and she has expressed interest in the Math Olympiad program which she learned about through a summer program she attended.

In this example, the student is in grade 2 and is 7 years and 8 months old. She has a raw score of 32 on the General Intellectual Ability (GIA) subscale and a raw score of 38 on the Mathematics subscale. Converting these raw scores into standard scores yields a 121 (92^{nd} percentile) on the GIA and 136 (99^{th} percentile) on the Mathematics subscales. In my example, I include the standard error of measurement (*SEM*), which is based on the reliability of the test. The *SEM* for the GIA subscale is 4 and for the Mathematics subscale is 3. In this case, a 68% confidence interval (CI) was used. This interval is 118 to 125 for the GIA subscale and 133 to 139 for the Mathematics subscale. In addition to the SIGS-2, the district uses two subtests from the SAGES-3: K-3: the Nonverbal Reasoning and the Mathematics/Science subtests. The SAGES-3 also provides raw scores, percentile ranks, and standard scores. The student in my example received a raw score of 21 on the Mathematics/Science subtest and 20 on the Nonverbal Reasoning subtest. These raw scores convert to index scores of 124 (95^{th} percentile) and 133 (99^{th}

percentile) respectively. With the standard error of measurement, the 68% confidence interval is 120 to 128 for the Nonverbal Reasoning subtest and 128 to 138 for the Mathematics/Science subtest.

The placement committee then should compare the standard score and the 68% confidence interval around the standard score to the target score which the district has set and check criterion met accordingly. Given her scores on multiple measures and the qualitative information, this student should be included in the gifted program.

Universal Screening

In the current NAGC State of the States Report (Rinn et al., 2020) several questions focused on universal screening. Universal screening (see Chapter 3) is when all students are administered at least one formal assessment as the first step in the identification of students who are gifted and talented. Of the 50 respondents in the survey, 9 indicated that universal screening was used for referral to identification, 9 indicated that it is used for identification, 32 stated it is not required, and 25 indicated it is up to the Local Education Agency (respondents could choose more than one response). The report did not specify which instruments districts used as universal screeners. Universal screening, however, can be particularly useful to identify certain groups of students in gifted and talented programs such as English language learners, ethnic and racial minority, low socioeconomic status, and twice-exceptional students (see Card & Guiliano, 2016). As discussed earlier, these students are often underrepresented in gifted and talented programs. Psychometrically sound rating scales are particularly useful as universal screeners because of their ease of administration. In addition, using rating scales that can be administered and scored online as universal screeners minimizes scoring mistakes.

Local Norms

Schools with students whose demographics are not reflective of the U.S. Census should consider using local norms. Local norms compare students' scores on a test or rating scale to peers using a local (e.g., peers in the same school building) sample rather than a national sample. Using national norms provides information on which students in a school scored higher than a particular percent of other students in the nation, while using local norms provides information on which students in a school scored higher than a particular percent of their peers in the same grade and school (Peters et al., 2019). It is important to note that it is not necessary to rate all students in a school to calculate local norms. Rather, the school can rate a stratified, random sample of its students and use the sample results to calculate the local norms. The strata should be chosen from the most relevant characteristics in the school. For example, if the school has a much higher proportion of students who are low-income than those in the U.S. Census, the school should sample low-income students in proportion to their representation in the school. Many of the rating scales described in the appendix include instructions to develop local norms. Some schools might choose to go beyond creating local norms and create group specific norms, which are calculated for a particular group of students such as low-income students.

Use of Psychometrically Sound Instruments

Schools should use rating scales that have sound psychometric properties and should consider setting the cutoff criterion low if using rating scales in the referral phase. Rating scales are often used in the referral phase and, if they score high enough, are given additional measures. If students score high enough on the additional measures, they are placed in the gifted program. When using referrals as the first phase of identification, it is critical to use high validity instruments whose scores are correlated with the scores from one or more confirmatory measures. All too often, referral cutoffs are set too high (i.e., in the gifted range), resulting in few false positives but many false negatives. False positives are when students are identified who should not be identified as gifted and false negatives are when students are not identified who should be identified as gifted. Furthermore, when referral cutoffs are equal or close to the confirmation test threshold (e.g., at the 90th or 95th percentile), the result is low sensitivity or many false negatives (McBee et al., 2016).

Because all assessments contain measurement error, it is imperative to set cut off scores in the referral phase lower than what typically occurs in many school districts. Setting the cut-off score depends on the type of gifted program the district provides and the effect an incorrect decision has on the student. For example, by using a more liberal cut off score of the 75% percentile, more students with potential gifts and talents will continue to be considered for program services.

Challenges Using Rating Scales

The biggest challenge of using rating scales to identify students as gifted is response bias (sometimes referred to as rater bias). Response bias means that respondents complete their ratings in ways that do not accurately reflect their true responses, typically because of certain teacher characteristics. These characteristics include experience variables, such as teaching experience and knowledge of giftedness; professional status such as teacher efficacy; or status variables such as students' ethnicity or income level (Mason et al., 2014). One type of response bias is response sets, which means that raters respond in a certain way to item formats (e.g., Likert scales) regardless of item content (Crocker & Algina, 1986). Response sets occur when examiners rate all items on the high (typically agree) or low (typically disagree) end or when examiners make differential interpretations of indefinite qualifiers such as often or some. Response bias negatively affects behavior rating scales' ability to classify students as needing intervention and/or service provision, in this case services in gifted programs (Mason et al., 2014).

Response bias frequently occurs when teachers overlook talent in students who do not fit the traditional definitions or do not possess the traditional characteristics of giftedness (Siegle & Powell, 2004). All too often this occurs in the referral phase. McBee (2006) found that teacher referrals and automatic referrals (i.e., students in Georgia are automatically referred for further testing if they score in the 90th percentile or higher on a standardized test), were far superior in accuracy than other referral sources. On the other hand, teacher referrals were less accurate for low income, Black, and Hispanic students than for high income or other races/ethnicities. This may be because of the subjective nature of the referral process or possible response bias. Using well developed and psychometrically sound rating scales during the referral process should alleviate some of this potential response bias. In addition, when teachers are trained in the characteristics of giftedness and, that these characteristics might manifest themselves in nontraditional ways, response bias is lessened. When using parent rating scales, educators should meet with parents who may have difficulty interpreting items on the rating scale.

Rating scale developers must use best practices when writing items and should describe this process in their technical manuals. In addition, experts should review rating scale items for accuracy, wording, ambiguity, and other technical flaws (Crocker & Algina, 1986). Examiners should read the technical manual to determine how items were constructed (see Price, 2017 for more information).

Recommendations

The discussion in this chapter leads to several recommendations that should be followed by educators responsible for the identification process of gifted students in a district. Rating scales, in either the referral or identification phases, should be used for their intended purposes. Sometimes districts use rating scales in a way that increases the probability of many false negatives. In other words, students who should be identified as gifted are not. Following the recommendations below will result in more accurate identification.

1. *Use rating scales that have good psychometric properties*. Educators should read the technical manual to ensure there are studies of reliability and validity and that these studies yield adequate findings. Some of the studies should be completed with students identified as gifted (e.g., reliability). The sample from which norms are created should reflect the U.S. Census in geographical region, sex, race, and ethnicity.
2. *Use rating scales that are based on observations of behavior that consist of a broad sampling of the most important characteristics of students who have talent*. The behaviors described on a rating scale must be observable in the setting in which the student is being observed. The behaviors should clearly represent what it means to be gifted in the area being rated.
3. *Train educators to use rating scales and to recognize characteristics of giftedness, especially in students who exhibit nontraditional behaviors*. Educators need to understand the characteristics they are looking for in students, some of which they may not be aware of as associated with gifted behavior such as "intense interest and curiosity about the world" or "asking many questions" or "offering creative or unusual responses to questions." Without training on the categories and items found on a rating scale, teachers may not feel comfortable scoring students at appropriate levels. Sometimes teachers have low expectations of low-income, culturally/linguistically diverse, and twice-exceptional students. Training teachers to recognize these nontraditional behaviors increases their awareness of these students.
4. *Ensure that rating scale scores are used for the purposes for which they are intended*. Purposes should be clearly spelled out in a rating scale's technical manual. It is the responsibility of the users to ensure that rating scales and other assessments are used correctly.
5. *Use rating scales with items that match the gifted programs for which students are selected*. It is critical to use scales that align with the program offered by the district. This means that not all the subscales on a rating scale will necessarily provide important information for identification of gifted students because the behaviors described do not match the district's program.

6. *Meet with and train parents who may have difficulty interpreting rating scale items.* This may be a result of linguistic differences or other reasons, but parents who have difficulty understanding what items mean will possibly not be able to complete the rating scale as intended.
7. *Use rating scales as universal screeners or set the cut-off scores lower than the "gifted" range when using them during the referral phase.* As discussed earlier, the consequences of setting the cut-off score too high increases the probability of many false positives. Setting the cut-off score lower and including the standard error of measurement increases the probability of correct identification.
8. *Create local norms when the demographics of the school are vastly different from the U.S. Census.* Local norms compare a student's performance with other students in the school or district while national norms compare students' performance of students in the U.S. Schools could also consider using group norms, which are created for a subgroup, such as low-income students in the school or the district.

Appendix C

Rating Scales to Assist in the Identification of Gifted Students

Several rating scales can be used to assist in identifying students as gifted, and these are described in this appendix. Most of the rating scales described in this appendix conducted several reliability and validity studies which are reported in their manuals. For brevity's sake, I present the results of the most common reliability and validity studies. For reliability, I report internal consistency and test-retest study results (and interrater if applicable) and for validity, I report criterion-prediction (i.e., correlations among the rating scales and other tests) study results.

The Scales for Identifying Gifted Students-Second Edition (SIGS-2; Ryser et al., 2021) consists of a School and Home Rating Scale, each of which have seven subscales: General Intellectual Ability, Language Arts, Mathematics, Science, Social Studies, Creativity, and Leadership. Each subscale has 10 items, which are rated using a 0 through 4 scale with 0 = Never and 4 = Much More. The rater is asked to compare the student to their grade-level peers of similar background and social status. The SIGS-2 is norm-referenced based on a national sample, although examiners are given the option and instructions for creating local norms. The SIGS-2 is available in both pencil/paper and online formats. Internal consistency reliability coefficients range from .81 to .97; test-retest reliability coefficients range from .61 to .93; and interrater reliability (teacher and parent) coefficients range from .43 to .59. Criterion- prediction coefficients ranged from .44 to .74, with one exception, which was the correlation between the SIGS-2 General Intellectual Ability subscale on the Home Rating Scale and the Torrance Tests of Creative Thinking Figural subscale. Several other validity studies were reported including results of a confirmatory factor analysis and a differential item functioning study.

The Universal Talented and Gifted Screener (UTAGS; McCallum & Bracken, 2018) consists of six subscales: Cognition, Creativity, Leadership, Literacy, Mathematics, and Science. The subscales can be used individually or can be combined into a composite score, which the authors call the General Aptitude Index. Teachers can rate students ages 5 through 17 using the UTAGS. Each subscale consists of 15 items, which are rated on a 1 to 5 scale with 1 = well below average to 5 = well above average. The UTAGS is norm referenced, although the authors provide information on how to create local norms. Internal consistency coefficients range from .98 to .99 and test-retest coefficients range from .84 to .96. Criterion-prediction coefficients range from .25 to .73.

The HOPE Teacher Rating Scale (Gentry et al., 2015) was designed to help identify and serve high-potential students from low-income families. The HOPE scale consists of 11 items rated on a 1 to 6 scale with 1 = Never to 6 = Always. The HOPE scale has two subscales Academic and Social with the Academic subscale consisting of six items and the Social subscale consisting of five items. Raters are asked to rate students as compared to other children who are similar in age, background, culture, and/or environment. The HOPE scale is not norm-referenced, but the manual provides instructions for creating local norms. Internal consistency reliability is .96 (Academic) and .92 (Social). Criterion- prediction validity coefficients range from .46 to .56. The authors also ran both an exploratory factor analysis and a confirmatory factor analysis.

The Gifted and Talented Evaluation Scales-Second Edition (GATES-2; Gilliam & Jerman, 2015) has five subscales: General Intellectual Ability, Academic Skills, Creativity, Leadership, and Artistic Talent. Each subscale consists of ten items that are rated on a 9-point scale divided into three ranges: 1–3, below average; 4–6 average; and 7–9 above average. The GATES-2 can be used to rate students ages 5 through 18 by individuals who know the student well. This is typically teachers, parents, or other professionals who have regular and sustained contact with the student. The GATES-2 is norm-referenced. Internal consistency coefficients range from .96 to .98, test-retest coefficients range from .81 to .88, and interrater coefficients range from .78 to .93. Criterion- prediction validity coefficients range from .08 to .92.

The Scales for Ratings of the Behavioral Characteristics of Superior Students-Third Edition (SRBCSS-III; Renzulli et al., 2010) is used to examine high potential and has fourteen subscales: Learning, Creativity, Motivation, Leadership, Artistic, Musical, Dramatics, Communication–Precision, Communication–Expressiveness, Planning, Mathematics, Reading, Technology, and Science. Subscales vary in length from four to 15 items which are rated on a 1 to 6 scale with 1 = Never and 6 = Always. The SRBCSS can be used with students ages 5 through 18. The SRBCSS is not norm-referenced; however, the manual provides instructions for creating local norms. Internal consistency coefficients range from .84 to .97, test-retest was not reported, and interrater reliability coefficients range from .50 to .65. Criterion- prediction coefficients range from .40 to .95.

The Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003) consist of two forms: the GRS–Preschool/Kindergarten Form for students ages 4–0 through 6–11 years and the GRS–School Form for students ages 6–0 through 13–11 years. The GRS–P has five subscales: intellectual, academic readiness, motivation, creativity, and artistic talent and the GRS–S has six subscales: intellectual, academic, motivation, creativity, leadership, and artistic talent. All subscales have 12 items, which are rated on a 9-point scale divided into the following three ranges: 1–3, below average; 4–6 average; and 7–9 above average. The GRS is norm-referenced and based on a national sample. Both forms of the Gifted Rating Scales are norm-referenced. Internal consistency coefficients were equal to or greater than .97 for each form. Test-retest correlations for the GRS-S ranged from .83 to .97. Interrater reliability coefficients for the GRS-P ranged from .70 to .84 and for the GRS-S, from .70 to .79 for ages 6:0–9:11 and from .64 to .75 for ages 10.0–13.11.

References

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly*, *60*(2), 81–101. https://doi.org/10.1177/0016986216634438

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Callahan, C. M., Moon, T. R., & Oh, S. (2013). *Status of elementary gifted programs*. Charlottesville, VA: National Research Center on the Gifted and Talented University of Virginia. www.nagc.org/sites/default/files/key%20reports/ELEM%20school%20GT%20Survey%20Report.pdf.

Card, D. & Guiliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(48), 13678–13683. https://doi.org/10.1073/pnas.1605043113

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.

Ford, D. Y. (1998) The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education*, *32*(1), 4–14.

Frasier, M. M., Hunsaker, S. L., Lee, J., Finley, V. S., Garcia, J. H., Martin, D., & Frank, E. (1995). *An exploratory study of the effectiveness of the Staff Development Model and the Research-based Assessment Plan in improving the identification of gifted economically disadvantaged students*. University of Connecticut, National Research Center on the Gifted and Talented.

Gentry, M., & Mann, R. L. (2008). *Total school cluster grouping and differentiation: A comprehensive, research-based plan for raising student achievement and improving teacher practices*. Creative Learning Press.

Gentry, M., Pereira, N., Peters, S. J., McIntosh, J. S., & Fugate, C. M. (2015). *HOPE Teacher Rating Scale*. Prufrock Press.

Gilliam, J. E., & Jerman, O. (2015). *Gifted and Talented Evaluation Scales-Second Edition*. PRO-ED.

Hunsaker, S. L., Finley, V. S., & Frank, E. L. (1997). An analysis of teacher nominations and student performance in gifted programs. *Gifted Child Quarterly*, *41*(2), 19–24. https://doi.org/10.1177/001698629704100203

Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment*, *20*(4), 322–336. https://doi.org/10.1177/073428290202000401

Johnsen, S. K. (2018). Making decisions about placement. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 117–149). Prufrock Press.

Johnsen, S. K. & Corn, A. L. (2019). *Screening Assessment for Gifted Elementary and Middle School Students-Third Edition*. PRO-ED.

Katz, I., & Rudolph, C. W. (2018). Rating scales. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1375–1377). SAGE Publications, Inc. www.doi.org/10.4135/9781506326139

Klein, B., & Fodor, S. (2019). TalentTiles: A new descriptive talent identification instrument based on teachers' ratings. In R. F. Subotnik, S. G. Assouline, P. Olszewski-Kubilius, H. Stoeger, & A. Ziegler (Eds.), *The future of research in talent development: Promising trends, evidence, and implications of innovative scholarship for policy and practice*. *New Directions for Child and Adolescent Development*, *168*, 11–25. https://onlinelibrary.wiley.com/toc/15348687/2019/2019/168

Likert, R (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 5–55.

Mason, B. A., Gunersel, A. B., & Ney, E. A. (2014). Cultural and ethnic bias in teacher ratings of behavior: A criterion-focused review. *Psychology in the Schools*, *51*(10), 1017–1030. https://doi.org/10.1002/pits.21800

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education*, *17*(2), 103–111. https://doi.org/10.4219/jsge-2006-686

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

McCallum, R. S., & Bracken, B. A. (2018). *Universal Talented and Gifted Screener*. Prufrock Press.

Moore, J. L., Ford, D., & Miler, H. R. (2005). Underachievement among gifted students of color: Implications for educators. *Theory into Practice*, *44*(2), 167–177. https://doi.org/10.1207/s15430421tip4402_11

National Association for Gifted Children. (2019). *Pre-K-Grade 12 Gifted Education Programming Standards*. NAGC. www.nagc.org/resources-publications/resources/national-standards-gifted-and-talented-education/pre-k-grade-12#:~:text=The%202019%20Pre%2DK%2DGrade, high%20quality%20programs%20and%20services.

Peters, S. J., & Gentry, M. (2012). Group-specific norms and teacher-rating scales: Implications for underrepresentation. *Journal of Advanced Academics*, *23*(2), 125–144. https://doi.org/10.1177/1932202X12438717

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who gets served in gifted education? Demographic representation and a call for action. *Gifted Child Quarterly*, *63*(4), 273–287. https://doi.org/10.1177/0016986219833738

Pfeiffer, S. & Jarosewich, T. (2003). *Gifted Rating Scales*. Multi-Health Systems Inc.

Price, L. R. (2017). *Psychometric methods: Theory into practice*. The Guilford Press.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., Gavin, M. K., Reis, S. M., Siegle, D., Sytsma Reed, R. E. (2010). *Scales for Rating the Behavioral Characteristics of Superior Students*. Prufrock Press.

Renzulli, J. S., & Westberg, K. (1991). *Scales for rating students' participation in the local gifted education program*. University of Connecticut.

Ricciardi, C., Haag-Wolf, A., & Winsler, A. (2020). Factors associated with gifted identification for ethnically diverse children in poverty. *Gifted Child Quarterly*, *64*(4), 243–258. https://doi.org/10.1177/0016986220937685

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018–2019 State of the States in Gifted Education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. www.nagc.org/2018-2019-state-states-gifted-education

Rothenbusch, S., Voss, T., Golle, J., & Zettler, I. (2018). Linking teacher and parent ratings of teacher-nominated gifted elementary school students to each other and to school grades. *Gifted Child Quarterly*, *62*(2), 230–250. https://doi.org/10.1177/0016986217752100

Ryser, G. R. (2018a). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 33–57). Prufrock Press.

Ryser, G. R. (2018b). Fairness in testing and nonbiased assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 59–72). Prufrock Press.

Ryser, G. R., McConnell, K., Sanguras, L. Y., & Kettler, T. (2021). *Scales for identifying Gifted Students-Second Edition*. Prufrock Press.

Siegle, D., & Powell, T. (2004). Exploring teacher biases when nominating students for gifted programs. *Gifted Child Quarterly*, *48*(1), 21–29. https://doi.org/10.1177/001698620404800103

Westberg, K. (2012). Using teacher rating scales in the identification of students for gifted and talented services. In S. L. Hunsaker (Ed.), Identification: The theory and practice of identifying students for gifted and talented services (pp. 363–379). Prufrock Press.

Worrell, F. C., & Erwin, J. O. (2011). Best practice in identifying students for gifted and talented education programs. *Journal for Applied School Psychology*, *27*(40), 319–340. https://doi.org /10.1080/15377903.2011.615817

CHAPTER

6

The Inclusion of Underrepresented Populations in Gifted Programs

Joyce VanTassel-Baska and Susan K. Johnsen

Introduction

Students of color, students from poverty, English language learners (ELs), and twice-exceptional students frequently differ from their peers in the general gifted population in specific ways and are therefore often overlooked in the identification process. Researchers suggest that these students are not considered for gifted programs and services because of the access to opportunities, implicit biases, use of inappropriate assessments, and a curriculum that does not play to their strengths (see Angelelli et al., 2002; Mun et al., 2016; Olszewski-Kubilius & Clarenbach, 2014; Weinfeld et al., 2021). For example, learners from low-income backgrounds often have depressed ability and aptitude scores, based on limited exposure to quality stimulation and early education. Students of color may not be referred because of implicit bias and racism. Students whose heritage language is not English may be left out of gifted programs until they reach an appropriate linguistic competency level. Even then, they may not be systematically considered because they lack important conceptual and cultural understandings in English. Students with gifts and disabilities (2E) may not qualify for programs that require high-level functioning in all areas of learning. It is difficult for 2E students to be identified for gifted and/or special education programs using traditional assessments, as they are likely to develop compensatory strategies that mask either the disability or the talent (Kirk et al., 2014). Educators may see test scores that place the student within the average range, because a student's giftedness is tempered by opportunities, access, or learning disabilities; at the same time, a student who should be identified for special services, for learning problems, for example, will be denied because his or her giftedness lifts test scores and academic achievement beyond the level required for services (Weinfeld et al., 2021). Some researchers have focused on identification techniques that appear to enhance participation of low-income learners and minority groups (Lakin & Lohman, 2011), yet patterns of underrepresentation of learners with special needs have persisted in identification practices across the United States for decades (McClain & Pfeiffer, 2012). If identified, many of these students with special needs underperform when faced with rigid curriculum structures that require narrowly defined behaviors

DOI: 10.4324/9781003285991-8

and responses (VanTassel-Baska et al., 2009) or skill sets not acquired at earlier ages through opportunities external to school (Burney & Beilke, 2008). Each special population faces challenges unique to their group.

Students with Gifts and Disabilities

Students with both gifts and disabilities are considered twice-exceptional (2E). The National Twice-Exceptional Community of Practice (2e CoP) developed this common definition for twice exceptional students (Baldwin et al., 2015):

> Twice exceptional individuals evidence exceptional ability and disability, which results in a unique set of circumstances. Their exceptional ability may dominate, hiding their disability; their disability may dominate, hiding their exceptional ability; each may mask the other so that neither is recognized or addressed. (p. 212)

Of the 13 disability categories identified under the Individuals with Disabilities Education Act (2004), all but one (mental retardation) could exist in gifted students, yet only three of the twice-exceptionality areas have been investigated: gifted students with specific learning disabilities, gifted students with attention deficit disorders with or without hyperactivity, and gifted students with autism spectrum disorder (Foley-Nicpon et al., 2011). Researchers have also found that comorbidity is a common problem for these students, suggesting that learning disabilities often pair with Attention Deficit/Hyperactivity Disorder (ADHD), autism spectrum disorders (ASD), depression, anxiety disorders, and various other complicating variables, such as minority status and low income, that can make identification even more challenging (Olszewski- Kubilius & Clarenbach, 2014; Weinfeld et al., 2021). Often overlooked in the identification of twice-exceptional students are those with physical disabilities that may make it challenging to identify and to serve in advanced programming. These are students who are deaf, visually impaired, those with cerebral palsy, and those with other physically limiting conditions.

Characteristics of twice exceptional students will vary then not only with the child within his or her talent domain but also with the disability. For example, children who have writing disabilities may be highly verbal but not be able to spell or write cohesive paragraphs. On the other hand, children who are deaf may have weaker oral communication skills but have a superior memory and be able to solve problems. These students' challenges often mask their gifts, which result in negative perceptions by adults, peers, and self. They may not be viewed as needing either gifted education or special education because they have some characteristics of both groups.

To date, it is not clear how many students are twice-exceptional. Researchers provide conservative estimates of 2E students ranging from 2% to 5% of children with disabilities (Nielsen, 2002) or up to 7% of students with disabilities (Trail, 2011). Barnard-Brak et al., (2015) examined the incidence of giftedness using a nationally-representative dataset and estimated that 9.1% of children with identified disabilities scored above the 90th percentile on standardized achievement assessments. The researchers indicated this percentage underestimates the population because disabilities often lower achievement scores, concluding that 2E students are "definitely underidentified" (p. 78).

Psychologists, special educators, and classroom teachers who are less familiar with twice-exceptionality than gifted specialists frequently do not refer these students for specialized programs and services (Foley-Nicpon et al., 2013). Along with educator misconceptions and limited professional learning, tests are not sensitive to 2E students, which may even lead to an overidentification of students with ADHD (Wood, 2012). For example, in a survey, the Belin-Blank Center identified 14 gifted students with a specific learning disability and found that one was receiving special education services only, and eight were receiving gifted services only (Assouline et al., 2010). Services are limited when students are not identified. These students, although gifted, may experience great difficulty in negotiating learning pathways due to their deficits in learning, attention, and socialization behaviors (Foley-Nicpon et al., 2011).

Gifted English Language Learners (ELs)

Gifted ELs are those who are "learning English as an additional language" (National Association for Gifted Children [NAGC], 2011). According to the National Center for Education Statistics (2017), Spanish was the home language of 3.7 million ELs in 2014–2015, representing 77.1% of all ELs and 7.6% of all public K–12 students. Arabic, Chinese, and Vietnamese were the next most common home languages (spoken by approximately 109,000; 104,000; and 85,300 students, respectively). In 2014–2015, a greater percentage of public school students in lower grades than in upper grades were ELs. For example, 16.7% of kindergarteners were ELs, compared to 7.8% of sixth graders and 6.5% of eighth graders. Among 12th graders, only 4.1% of students were ELs. This pattern is driven, in part, by students who are identified as ELs when they enter elementary school but obtain English language proficiency before reaching upper grades. The issue of gifted students who are ELs is more critical than at the elementary level, where they may be left out of gifted programs until they reach an appropriate linguistic competency level. Even then, they may not be systematically considered because of cultural differences, teacher perceptions, parent advocacy, and verbally-laden instruments.

These students bring diverse cultural and linguistic backgrounds into the educational arena, but for many of them, the struggle with the language and culture of school feels overwhelming. It is not surprising, then, that ELs are often ignored for inclusion in gifted programs. Their lack of inclusion also relates to teacher perceptions of seemingly insurmountable problems with language acquisition. These students' parents usually face as many or more language barriers, meaning that parents who want to advocate often cannot. These parents, then, through no fault of their own, are silent on the issue of their child's educational needs (Angelelli et al., 2002; Arias & Morillo-Campbell, 2008). In addition to these challenges, verbally-laden assessments preclude their identification. Language becomes a barrier and not a window on the intellect (McCallum, 2017). With 200 or more languages, test translations and subsequent norming and validation can become costly and time-consuming. One possible solution is the use of nonverbal tests; however, the user must be cautious to ensure that the instrument does not incorporate language in its directions nor require language in its responses.

Gifted Students of Color

Students of color is a term used for those who identify as Black or African-American, Hispanic or Latinx, Asian, American Indian or Native American, and/or multiple racial identities. According to the 2019 United States Census Bureau 60.1% of the population identify as White alone, 18.5% as Latinx, 13.4% as Black or African American, 5.9% as Asian, 1.3% as American Indian and Alaska Native, and the remainder as two or more races. While the U.S. and school populations are becoming more diverse, most students of color are underrepresented in gifted education programs. For example, the National Center for Education Statistics (2017) reported that 6.7% of public school students are enrolled in gifted and talented programs but only 4.9% are Hispanic or Latinx and 4.3% are Black or African American. The only group that is consistently overrepresented are Asian (13.3%).

Implicit bias and racism have been found to negatively impact identification and educational opportunities (Carnevale et al., 2019; Ford et al., 2011; McBee, 2010). Educators may see these students as reflecting group stereotypes without fully recognizing their abilities. Consequently, teachers may be less likely to refer students for gifted programs who are less acculturated and have differences in verbal ability (Hamilton et al., 2018; Plucker & Peters, 2018). VanTassel-Baska et al. (2009) found that students from different minority backgrounds preferred different modes of expression other than verbal to communicate learning. African Americans were likely to enjoy the expressive arts as a mode of talent display, often using nonverbal approaches to convey their ideas in music, dance, and the visual arts. Hispanic students were similar in that regard, perhaps due to language issues. Asian students also preferred nonverbal media to present their abilities, often in math and science-related areas. Moreover, some ethnically diverse students do not feel that they belong in gifted and talented education programs and may feel they have to choose between high academic achievement and being a genuine member of their racial/ethnic group (Worrell, 2007; Worrell & Dixson, 2018). Identifying and retaining these students requires attention not only to assessments but also cultural and psychosocial variables (Dixson & Stevens, 2018).

Black, Hispanic, and American Indian students come from households that are on average less affluent than the households of peers and tend to be concentrated in high poverty schools (Hamilton et al., 2018). Over half of Hispanic students attend schools with more than 75% of the students on lunch assistance, yet the pattern for White and Asian students is very different, with 71% and 65%, respectively, attending schools with 50% or fewer of the students qualifying for lunch aid (Plucker & Peters, 2016). This concentrated poverty coupled with minority designations creates additional issues in providing educational equity.

Students from Poverty

Over the past generation, the percent of K-12 students qualifying for free or reduced-price lunch programs, one of the most widely-used metrics of low-income status, has substantially increased (Plucker & Peters, 2018). For the 2018–2019 school year, 52.3% of students qualified for these programs, meaning that over half of our public school students live in households whose income is 1.85 times the poverty level or less (National Center for Education Statistics, 2019). For a family of four, this means an income of slightly less than $50,000 a year. That rate is up from 49.6% in 2011–2012, continuing the trend since the last economic crisis of a roughly 1.5% annual increase.

Being poor is also a status, despite the American dream, that all but perhaps 5–7% of Americans are likely to remain in due to the insidious relationship of educational attainment and job accessibility. Sennett and Cobb (1972) and more recently Chetty et al. (2014) confirm the difficulty in moving up from the lower class in our society, with fewer than 7.5% of the population moving out of poverty over the last ten years. These data suggest that social mores, neighborhood, parental education and career, and the prevalence of divorce and child-bearing all collude to make such upward mobility unlikely.

Students from impoverished backgrounds are at greater risk for a host of social-emotional problems, including lower levels of motivation when compared to children who do not come from such backgrounds (Ambrose, 2013). Oftentimes, the risks for social-emotional problems come from related special challenges for students living in poverty, including higher rates of disabilities, teenage mothers, absent fathers, lower motivational levels, parents without resources, health problems, concerns about safety and daily survival, and increased risk of homelessness (Duncan & Murname, 2011; Stormont et al., 2001). Other negative effects of poverty are mediated through family and community dynamics such as a less enriching home environment, authoritarian and controlling parenting, and fewer community institutions that provide support (Engle & Black, 2008). For example, a study, using data from the National Longitudinal Study of Youth, demonstrated that families above the poverty line were more likely to engage in cognitively enhancing activities with their children than were families below the poverty line (Bradley et al., 1994). Likewise, higher income and more educated families were more likely to read to their children before Age 5.

Research has documented specifically the adverse effects of poverty on learning and achievement outcomes (Lacour & Tissington, 2011; Moore et al., 2009; Ellis et al., 2017)). Poverty affects educational opportunity (e.g., school choice, availability of early childhood education) and school quality (e.g., fewer advanced classes, less experienced teachers, higher teacher turnover (Aaronson et al., 2007; Rivkin et al., 2005). It limits the financial and person power resources available within the family to support learning beyond the school day such as trips to museums, participation in cultural events, and extra-curricular or summer programs (Snellman et al., 2015). Poverty can also impact students' aspirations, including whether or not they view a future for themselves that includes higher education and professional careers by limiting access to peer and/or adult mentors and professionals who can assist with educational paths and career development (Snellman et al., 2015). The conditions of poverty continue to impact achievement throughout the school years in all subjects and areas of learning (National Center for Education Statistics, 2013). A national report observes, "Children born into low-income families face more challenges throughout their lives than children that are born into families with higher incomes. These challenges typically lead to poorer outcomes throughout the life span" (National Center for Education Statistics, 2013). Persistent poverty continues to cause achievement to deteriorate for these learners as seen in both state data and national longitudinal studies (Michelmore & Dynarski, 2016).

Yet early intervention programs have shown clear benefits even as the United States has failed to provide quality preschool and primary programs for the poor on a widespread basis. Reports of intensive early intervention with poor children have shown effect sizes of 0.5–0.75 for the Perry Preschool Project, the Abecedarian Project, and the Chicago Child–Parent Centers.

The Perry Preschool Program and the Abecedarian studies were randomized controlled trials of early educational programs that targeted low-income children and showed benefits that extended beyond formal school years into adulthood. The Perry Program began when children were ages 3 and 4 and provided intensive preschool education and home visits for children in poverty. The Abecedarian Project began during the first year of life and provided intensive services to poor and primarily African American mothers and children for 5 years. Both were intensive, high-quality efficacy studies. The Chicago program offered comprehensive services to low income families of students at K-3 levels. Longitudinal follow-up showed that children in the intervention were more likely to graduate from high school, attend college, have fulltime employment, and be good adult citizens in comparison to those not enrolled (Engle & Black, 2008).

Within the gifted population, this group of learners is often overlooked in favor of finding underrepresented students of color, especially if they constitute a plurality within the school district. While no studies have documented this point specifically, the anecdotal evidence suggests that it occurs with some frequency. An analysis of intervention studies of these students suggest that the focus is on the overlap of underrepresented groups in gifted programs, not on the distinctive issues associated with each group. It may be fair to say that there is still a distinct group of students from all groups in the United States whose needs require attention for both identification and programming (Ambrose, 2013). Hamilton et al. (2018) have found that students in low-income schools have lower identification rates, and students from low-income backgrounds in low-income schools are the least likely to be identified. Poverty, therefore, brings substantial risks for educational attainment and achievement in all areas of education.

Interrelated Factors Affecting Gifted Identification within Underrepresented Groups

It is not uncommon for students from underrepresented groups to be dually labelled, to be poor and of color, to be twice exceptional with co-morbid conditions, to be EL and poor. The intersection of English language proficiency and poverty within the context of gifted programs is similar in many ways to the intersection of race and poverty, with the primary similarity being a compounding of adversities. For instance, ELs from low income backgrounds have to navigate being identified as academically gifted in a system that is specifically tailored towards students with a different background (i.e., English native speakers) as well as contend with the multitude of challenges associated with trying to develop one's gifted academic potential with limited resources (Abedi, 2002; Lohman et al., 2008; Olszewski-Kubilius & Corwith, 2018). Also, ELs from low income backgrounds encounter pervasive negative perceptions about their group that work against them being identified as academically gifted and subsequently receiving services (see Pettit, 2011 for review). Langley (2017) has noted that identification is thwarted not just by language issues but also teacher views of deficits in particular aspects of communication demonstrated by ELs in classroom contexts. Studies have continued to see poverty, along with minority status, English as a second language, and single parent status as the critical variables contributing to lack of achievement or "disadvantagement" in the school system (Pallas et al., 1989; Reardon & Portilla, 2016). In sum, it appears that despite widening income inequality, increasing income segregation, and growing disparities in parental spending on children, disparities in school readiness narrowed from 1998 to 2010.

Research has also demonstrated that both race and poverty are large factors in students' identification as academically gifted with the relative importance of each varying depending on the context (Dixson, 2020; Elhoweris et al., 2005; McBee, 2010; Peters & Engerrand, 2016). Compared to rural settings, cities provide the context for more stimulation, more chances for meeting intellectual peers, and more access to resources for learning and applying skills to various talent areas. Data also suggest that cities have often been viewed as the contexts for high-level talent development, while rural areas have been viewed as less promising contexts for nurturing achievement and productivity in fields of endeavor (Howley & Showalter, 2015; Stambaugh & Wood, 2015), even though poverty is an issue in each type of demographic location. The farther students are from major cities and metropolitan areas, the less access they have to the knowledge, experiences, and people that can aid their talent development (Dixson, 2020). In a national study of elementary school gifted offerings, Callahan et al. (2013) found that rural school districts were significantly less likely to offer full-time gifted programing for their academically gifted students (13%) compared to urban (20.2%) and suburban (21.2%) school districts. Kettler et al. (2016) examined specific academic offerings of more than 1,000 schools and found that students schooled in rural contexts had less access to advanced academic programing than students being schooled in all other contexts, despite exhibiting comparably positive performance on advanced academic outcomes. Similarly, Mann et al. (2017) found that rural students had less access to Advanced Placement (AP) courses overall (urban-95%, suburban-92%, and rural-73%) and within STEM subjects (urban-88%, suburban-93%, and rural-62%) compared to their urban and suburban counterparts.

Effects of Limited Access to Advanced Programming

Educational achievement gaps between groups of children in the U.S. have been a major focus of policy and reform efforts (e.g., Every Student Succeeds Act and No Child Left Behind). Reardon (2011), using data from 19 nationally representative studies, examined the history of achievement gaps in the U.S. and concluded that the income gap in educational achievement, defined as the difference between children whose families were at the 90th versus the 10th percentile of family income, has widened and is now twice as large as the Black-White achievement gap. Therefore, similar effects can exist regardless of income if access and opportunity are not provided.

The achievement gap between lower and higher income children is present at school entry and persists as children progress through school. Carnevale et al. (2019) showed that the disparities on these factors between higher and lower income children has not changed since an earlier study in 2003 (see Barton & Coley, 2009). Disparities across race and socioeconomic status grow as students progress through school (Engle & Black, 2008; Olszewski-Kubilius et al., 2017), suggesting that small differences in resources and opportunities may lead to slight differences in academic ability at a given stage of development but compound over time (Ceci & Papierno, 2005).

The National Assessment of Educational Progress program (NAEP), often called the nation's report card, has documented the achievement of different groups of students in the U.S. since 1990. This research has shown large and persistent achievement gaps between students who do or do not qualify for the federal free and reduced lunch program (National Center for Education Statistics, 2017) since its inception. The NAEP data showed that the free and reduced lunch vs. the non-free and reduced lunch achievement gaps in mathematics and reading at grades 4 and 8 remained relatively stable from 2013 to 2019; however, gaps of 25 to 28 points in grades 4 and 8 in reading and gaps of 30 to 34 points in grades 4 and 8 in math were revealed in the most recent data (see NAEP Report Card at www.nationsreportcard.gov).

Moreover, higher achieving, lower income students lose ground in school (Xiang et al., 2011). In a Fordham Institute study, high performing low-income students (i.e., scoring at or above the 90th percentile) fell in achievement markedly from Grade 3 to Grade 6 in reading and/or math while higher income level high performing students did not experience such dips in achievement across the elementary years. Although these students did not fall below the 70th percentile, it hypothetically left them out of being eligible for many gifted programs.

Differences in exposure to language and vocabulary (Hart & Risley, 2003) and informal exposure to the natural world (Curran & Kellogg, 2016; Morgan et al., 2016) in children's early environments have been documented. Such disparities significantly impact children's readiness for school, put them behind at the start of school for their academic growth, persist and grow as children proceed through school (Fernald & Marchman, 2013; Morgan et al., 2016). These early and persistent achievement disparities have significant consequences for adulthood. A kindergartner who comes from a high socioeconomic status (SES) family and with test scores in the bottom 50% has a 7 in 10 chance of reaching high SES in adulthood while a kindergartner from a low SES family with test scores in the top half has only a 3 in 10 chance of being higher SES by the age of 25 (Carnevale et al., 2019).

Wyner et al. (2009), using data from three national longitudinal studies, found that:

1. High achieving, low SES secondary students have lower grades in their academic courses compared to higher SES students.
2. High achieving low SES students are less likely to take advanced mathematics courses in high school, nor take Advanced Placement or International Baccalaureate options that would better prepare them for the rigor of college. Those who do take AP tend to score lower than higher SES students.
3. High achieving low SES students attend less selective colleges than students more advantaged, but no more able, (21% vs.14%) are less likely to graduate from college (49% versus 77%) and are less likely to earn a graduate degree (29% vs. 47%).

These results suggest the persistence of disparities in achievement at all stages of preparation and development, affecting opportunities for college, career, and life paths.

Results of Survey and Interviews with Gifted Coordinators

The authors of this chapter surveyed and interviewed six coordinators of gifted programs who were involved with the issue of underrepresentation in their school district and actively were addressing it. The coordinators represented three states, worked in districts that ranged from small rural to large urban and had varying levels of expertise and experience in working with the issues involved in changing the identification model to accommodate more underrepresented learners. All of these six districts had high percentages of low income learners (i.e., greater than 50% across the district). All districts also had high percentages of students of color, depending on the subgroup. Most of the districts identified 13% of students who needed special education services (SPED).

Programmatic similarities were apparent across the districts. All but two districts used cluster grouping at the elementary level to serve gifted students with a trained teacher in gifted education at the helm of these classrooms. Teachers of the gifted also served as facilitators of the clusters, working with teachers, and providing pullout opportunities as needed. In one district students were served in a heterogeneous setting with teachers of the gifted providing pullout opportunities. Middle schools varied in respect to the degree of grouping provided for the gifted, ranging from no grouping to cluster grouping with high ability learners by subject area to special classes in selected areas of the curriculum. High school opportunities included Advanced Placement in all six districts and International Baccalaureate in three of them. Other secondary opportunities were individual mentorships, internships, and competitions. All of the districts reported concerns about the degree and nature of differentiation that was occurring within their programs and the training and support for such efforts in their districts.

Table 6.1 reports the major approaches these districts used to improve their identification model in respect to finding more underrepresented groups. All of these districts employed *multiple approaches* to address the problem that included the use of universal screening coupled with the use of multiple measures, coupled with enhanced professional learning for teachers and administrators. All of these districts were also trying to track the effectiveness of their approaches by examining *changes in proportionality* of these students in gifted programs. None indicated they were assessing other aspects of the new processes employed. Most of the districts were actively addressing *the use of early talent development programs for kindergarten-Grade 2* students, using gifted curriculum as a tool to uncover potential abilities in individual areas of learning. Coupled with such programs was also *active scouting for talent* in student populations within individual schools. Efforts to provide *tiers, both of selection and instruction*, was commonly supported by these coordinators. Several of them employed RtI approaches to instruction that allowed for

TABLE 6.1 Approaches to improving the number of underrepresented groups in gifted programs by six districts

Approach Employed	District #1	District #2	District #3	District #4	District #5	District #6
1. Universal screening	x	x	x	x	x	x
2. Early talent development program	x	x			x	x
3. Teachers as talent scouts	x					x
4. Use of subtest scores	x	x		x	x	x
5. Use of local norms	x	x	x	x	x	x
6. Professional learning	x	x	x	x	x	x
7. Dissemination of materials	x	x		x		
8. Grouping and differentiated instruction		x		x	x	x

Notes:

1. Universal Screening is assessing all children within a given class, grade, school campus or school district level on academic, ability, and/or social-emotional of indicators using qualitative and/or quantitative instruments.
2. An early talent development program engages students in elementary grades in accelerated and/or enriched programming taught by resource teachers in gifted education based on the student's strengths and needs.
3. General education and teachers who represent specialized programs (i.e., special education, English as a second language) are actively involved in searching for students who may show signs of potential as gifted.
4. The district uses subtest scores to look for discrepancies in performance that might uncover relative strengths and weaknesses of students and identify gifted students who have disabilities or language barriers.
5. Local norms are used rather than national norms when the demographics are not reflective if the district or campus student population. With local norms, students are more likely to be compared against other students with similar demographic characteristics.
6. Professional learning related to finding and serving gifted students from underrepresented populations is provided to all teachers in the district with specialists in gifted education receiving annual updates.
7. Dissemination of materials about gifted identification and programming is provided in multiple languages through online and personal contact.
8. Students are cluster grouped and receive differentiated instruction based on assessment information.

addressing gifted student needs at tiered levels of support; others used an in-district model for purposes of differentiating services. One district used an Individual Education Plan (IEP) as a tool to ensure an optimal match between the program and each gifted student. Changes in identification criteria for selecting students for programming were also commonly employed, including (a) lowering cutoff scores for eligibility on specific measures, (b) using local norms, including school-based norms for selection (c) using subtest scores as opposed to composite scores on ability tests and (d) focusing on the use of tools such as portfolios and products as evidence of readiness to participate in gifted programs.

Reaction to Changes in the Gifted Identification Model

All of the districts indicated that they enjoyed support for new practices to improve underrepresentation from their teachers and administrators within the district as well as in their state. In fact, teachers were actively involved in the processes adopted to promote equity. All three states had made changes (or were in the process of making changes) in state policy to accommodate efforts to increase the number of underrepresented students identified for gifted programs. These districts also made efforts to align state and district policies. Barriers to change more often came from the gap in knowledge of the issue rather than objection to specific changes in the identification process. Stakeholders in general at both teacher and administrator levels supported and even welcomed such changes.

In at least three of the districts (50%), it was clear that changes in personnel had impacted the new identification practices. In one district, portfolios were eliminated, even though data indicated that they helped include more underrepresented groups. In another, less communication about the processes was put in place. In a third, lack of knowledge of the overall gifted program by a new coordinator may have impeded the progress on these initiatives. In one of the other districts that did have continuity in leadership, the coordinator suggested that sustained leadership was a positive factor in moving this agenda forward.

Perceived Advantages and Disadvantages of Changing the Identification Process

Most of the benefits listed by coordinators of changing the system of identification rested on the belief that it produced greater equity in who was selected for gifted programs; less attention was paid to the actual changes that occurred in proportionality and what aspects of the new processes accounted for them, however. Disadvantages of the institution of these new processes that were cited included time and resources. Several coordinators mentioned that teachers and they themselves had to put in many extra hours in order to ensure that the identification system was implemented with fidelity. There was also a recognition that this was not a one-year issue, but rather one that would require multiple years of using multiple approaches. One coordinator put in place a five-year plan to acknowledge the complexity of addressing these issues; however, she did so just in respect to EL students, one of the underrepresented groups.

Stories of Successful Impact of Identification Policy Change

The coordinators were able to share success stories with us about the effects of the changes that were made. One coordinator shared a series of vignettes that illustrated the success of students who had been identified: "A third grade child whose native language is Amharic, was attending a high socioeconomic campus. She was afraid to be identified. She was not only EL but perhaps had a disability. However, she was identified and is now going to be accelerated, a service commensurate with her needs. We would also not have found Aydin, a fourth grader who sees himself as a scholar and a capable learner. He was identified as gifted through universal testing and says he loves taking on 'future adult problems' in his gifted class via Zoom with his teacher." Another coordinator talked of a child on the autistic spectrum who enrolled in a gifted class in 2nd grade: "We closed his gaps in writing. He grew in academic areas and was able to study his passions. He made friends. He taught the teacher and the other kids. The teacher had patience and gave him grace. The parent cried because her son found a (safe) place." Still another told of the power of the appeals process where an ethnically diverse student's grandparent was able to articulate his special abilities for the committee. Another story was related of a young girl who became interested in science due to her 6th grade project in the gifted program; she went on to be mentored in high school and produced high quality work, all not possible without the kinds of changes enacted in her district's identification process.

The responses from these coordinators, however, often did not specify the importance of different approaches for different groups. For example, addressing the language issue is critical in finding EL learners, while considering specific abilities through differences in subtest scores is critical for finding twice-exceptional learners. Often the issue of poverty per se was not addressed, even though the disparity in those identified from this demographic was the greatest in all of the districts. Problems uncovered through these interviews suggest several considerations that need to be implemented in establishing more effective identification procedures for underrepresented groups.

- It is important to tailor identification and services to the strengths, needs, and interests of underserved and under-resourced groups of students. Differences of "between group needs" as well as "within group (individual) needs" require attention.
- Multiple approaches, not just one, must be employed to make progress on identifying students from underrepresented groups. Universal screening, use of local norms, and early talent development curriculum appear to be the most common approaches employed in concert successfully.
- While follow-up data on proportionality of these populations in the gifted programs are collected, it is unclear how those data are used to improve approaches for future identification cycles and to optimize instructional services. Details of the outcomes from changing identification must become part of a written record to disseminate to school boards and advisory groups in the district to ensure fidelity of implementation, continuity, and improvement of efforts.
- There is a unitary focus in most of the districts on finding particular underrepresented groups. Typically, the focus is on these ethnic subgroups—Black and Latinx (EL). Other groups are more often ignored (e.g., students who are 2E and those from lower income backgrounds). While this leads to an additional problem of too narrow a view on who is underrepresented and why; it also leads to a false sense of progress on the issue if one group's participation becomes more proportional while others do not.
- Finally, it appears that there is not a sufficient emphasis on student outcomes related to learning as the centerpiece for making such changes. No district is

tracking the specific learning outcomes of these students once they are placed in more advanced programming. Without such data, the field is not likely to have a roadmap for program improvement nor for more effective approaches to finding underrepresented students in the future.

Recommended Principles for Identifying Underrepresented Students

The following principles summarize effective practices identified by experienced school personnel and the extant literature cited in this chapter. These principles are intended to assist practitioners in reducing disparities in the identification of underrepresented students in school systems and in showing educators and parents the interrelationships between identification and programming at the school level.

- Equal representation is desirable to enhance program diversity, which should be an equity goal for all gifted programs;
- The use of effective combinational approaches such as universal screening coupled with the use of local norms and the use of subtest rather than composite scores on cognitive and achievement tests together work to yield more students in underrepresented groups;
- The use of multiple tests and multiple sources to complete inventories is superior to a single test score and the judgment of just one educator;
- The use of current student performances and products adds a layer of authenticity to the process and identifies students' strengths, needs, and interests;
- The process of identification, whether in one or more stages, should be conducted by educators who have received professional learning in gifted education;
- The identification process needs to be continuously monitored to ensure that it is implemented with fidelity.

Conclusion

Finding and serving underrepresented groups as a part of the gifted program is a task that must be done. Recent evidence suggests that few districts have specific provisions for these students noted in their state plans nor report positive changes in status when new strategies are tried through such techniques as universal screening (McCoach, 2021). We cannot change a student's accident of birth, but we can positively effect a deliberate path toward greater fulfillment in life, partially defined by access, choice, sustainability of advanced education and a satisfying career. As Dixson (2020) has noted: "Given that academically gifted individuals with fully developed talent have been found to make disproportional contributions to society (Wai et al., 2019), a takeaway is how much further would society be if it made a more concerted effort to develop the academic talent of those without the resources and context to do it on their own?" (p. 20).

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1), 95–135. www.journals.uchicago.edu/doi/10.1086/508733

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, *8*(3), 231–257. https://doi.org/10.1207/S15326977EA0803_02

Ambrose, D. (2013). Socioeconomic inequality and giftedness: Suppression and distortion of high ability. *Roeper Review*, *35*(2), 81–92. https://doi.org/10.1080/02783193.2013.766960

Angelelli, C.V., Enright, K., & Valdés, G. (2002). *Developing the talents and abilities of linguistically gifted bilingual students: Guidelines for developing curriculum at the high school level* (RM02156). University of Connecticut, The National Center on the Gifted and Talented. https://nrcgt.uconn.edu/research-based_resources/angeenri/

Arias, M., & Morillo-Campbell, M. (2008). *Promoting ELL parental involvement: Challenges in contested times.* Arizona State University and University of Colorado-Boulder. https://files.eric.ed.gov/fulltext/ED506652.pdf

Assouline, S. G., Nicpon, M. F., & Whiteman, C. (2010). Cognitive and psychosocial characteristics of gifted students with written language disability. *Gifted Child Quarterly*, *54*(2), 102–115. https://doi.org/10.1177/0016986209355974

Baldwin, L., Baum, S., Pereles, D., & Hughes, C. (2015). Twice-exceptional learners: The journey toward a shared vision. *Gifted Child Today*, *38*(4), 206–214. https://doi.org/10.1177/1076217515597277

Barnard-Brak, L., Johnsen, S. K., Pond, A. P., & Wei, T. (2015). The incidence of potentially gifted students within a special education population. *Roeper Review*, *37*(2), 74–83. https://doi.org/10.1080/02783193.2015.1008661

Barton, P. E., & Coley, R. J. (2009). *Parsing the achievement gap II.* Policy Information Report. Education Testing Service. www.ets.org/Media/Research/pdf/PICPARSINGII.pdf

Bradley, R. H., Whiteside, L., Mundfrom, D. J., Casey, P. H., Kelleher, K. J., & Pope, S. K. (1994). Early indications of resilience and their relation to experiences in the home environments of low birthweight, premature children living in poverty. *Child Development. 65*(2), 346–360. https://doi.org/10.2307/1131388

Burney, V. H., & Beilke, J. R. (2008). The constraints of poverty on high achievement. *Journal for the Education of the Gifted*, *31*(3), 295–321. https://doi.org/10.4219/jed-2008-771

Callahan, C., Moon, T., & Oh, S. (2013). *Status of elementary gifted programs—2013.* Charlottesville, VA: University of Virginia National Research Center on the Gifted and Talented. www.nagc.org/sites/default/files/key%20reports/ELEM%20school%20GT%20Survey%20Report.pdf

Carnevale, A. P., Fasules, M. L., Quinn, M. C., & Campbell, K. P. (2019). *Born to win, schooled to lose: Why equally talented students don't get equal chances to be all they can be*. https://cew.georgetown.edu/cew-reports/schooled2lose/

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149–160. https://doi.org/10.1037/0003-066X60.2.149

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *The American Economic Review*, *104*(5), 141–147. https://doi.org/10.1257/aer/104.5.141

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade, *Educational Researcher*, *45*(5), 273–282. https://doi.org/10.3102/0013189X16656611

Dixson, D. (2020) Intersections of culture, context, and race with poverty: Implications for services for low-income gifted learners. In T. Stambaugh & P. Olszewski-Kubilius (Eds.), *Unlocking potential in children from low income households* (pp. 27–46). Prufrock Press.

Dixson, D. D., & Stevens, D. (2018). A potential avenue for academic success: Hope predicts and achievement-oriented psychosocial profile in African American adolescents. *Journal of Black Psychology*, *44*(6), 532–561. https://doi.org/10.1177/009579841880206

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, school and children's life chances, Executive Summary*. Russell Sage Foundation.

Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education*, *26*(1), 25–31. https://doi.org/10.1177/07419325050260010401

Ellis, B. J., Bianchi, J. M., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science, 12*(4), 561–587. https://doi.org/10.1177/1745691617693054

Fernald, A., & Marchman, P. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://10.1111/desc.12019

Engle, P. L., & Black, M. M. (2008). The effect of poverty on child development and educational outcomes. *Annals of the New York Academy of Sciences*, *1136*(1), 243–256. https://doi.org/10.1196/annals.1425.023

Foley-Nicpon, M. F., Allmon, A., Sieck, B., & Stinson, R. D. (2011). Empirical investigation of twice-exceptionality: Where have we been and where are we going? *Gifted Child Quarterly*, *55*(1), 3–17. https://doi.org/10.1177/0016986210382575

Foley-Nicpon, M., Assouline, S. G., & Colangelo, N. (2013). Twice-exceptional learners: Who needs to know what? *Gifted Child Quarterly*, *57*(3), 169–180. https://doi.org/10.1177/0016986213490021

Ford, D. Y., Moore III, J. L., & Scott, M. T. (2011). Key theories and frameworks for improving the recruitment and retention of African American students in gifted education. *The Journal of Negro Education*, *80*(3), 239–253. www.jstor.org/stable/41341131

Hamilton, R., McCoach, D. B., Tutwiler, M. S., Siegle, D., Gubbins, E. J., Callahan, C. Brodersen, A. V., & Mun, R. U. (2018). Disentangling the roles of institutional and individual poverty in the identification of gifted students. *Gifted Child Quarterly*, *62*(1), 6–24. https://doi.org/10.1177/0016986217738053

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by Age 3. *American Educator*, 27(1), 4–9. https://eric.ed.gov/?id=EJ672461

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108–446, 118 Stat. 2647 (2004).https://ies.ed.gov/ncser/pdf/pl108-446.pdf

Kettler, T., Puryear, J. S., & Mullet, D. R. (2016). Defining rural in gifted education research: Methodological challenges and paths forward. *Journal of Advanced Academics*, *27*(4), 245–265. https://doi.org/10.1177/1932202X16656896

Kirk, S., Gallagher, J. J., & Coleman, M. R. (2014). *Educating exceptional children* (14th ed.). Cengage Learning.

Lacour, M., & Tissington, L. D. (2011). The effects of poverty on academic achievement. *Education Research and Reviews*, *6*(7), 522–527. www.academicjournals.org/ERR

Lakin, J. M., & Lohman, D. F. (2011). The predictive accuracy of verbal, quantitative, and nonverbal reasoning tests: consequences for talent identification and program diversity. *Journal for the Education of the Gifted*, *34*(4), 595–623. https://doi.org/10.1177/016235321103400404

Lohman, D. F. (2005). *Identifying Academically Talented Minority Students,* National Research Center for the Gifted and Talented.

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-Language Learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 275–296. https://doi.org/10.1177/0016986208321808

Mann, S., Sponsler, B., Welch, M., & Wyatt, J. (2017, August). *Advanced placement access and success: How do rural schools stack up?* www.ecs.org/wp-content/uploads/Advanced-Placement-Access-and-Success-How-do-rural-schools-stack-up.pdf

McBee, M. (2010). Examining the probability of identification for gifted programs for students in Georgia elementary schools: A multilevel path analysis study. *Gifted Child Quarterly*, *54*(4), 283–297. https://doi.org/10.1177/0016986210377927

McCallum, R. S. (Ed.). (2017). *Handbook of nonverbal assessment* (2nd ed.). Springer.

McClain, M-C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*, *28*(1), 59–88. https://doi.org/10.1080/15377903.2012.643757

McCoach, B. (2021, March) *Overview of what gifted programs look like*. Online presentation at National Center for Research on Gifted Education Conference. https://ncrge.uconn.edu/recommendations-for-improving-gifted-education/

Michelmore, K. & Dynarski, S. (2017) The gap within the gap: Using longitudinal data to understand income differences in educational outcomes, *AERA Open*, *3* (1), 1–12. https://doi.org/10.1177/2332858417692958

Moore, K. A., Redd, Z., Burkhauser, M., Mbwana, K., & Collins, A. (2009, April). Children in poverty: Trends, consequences, and policy options. *Child Trends Research Brief*, 1–12. www.childtrends.org/wp-content/uploads/2013/11/2009-11ChildreninPoverty.pdf

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*(1), 18–35. https://doi.org/10.3102/0013189X16633182

Mun, R. U., Langley, S. D., Ware, S., Gubbins, E. J., Siegle, C. M., McCoach, D. B., & Hamilton, R. (2016). *Effective practices for identifying and serving English learners in gifted education: A systematic review of the literature*. National Center for Research on Gifted Education University of Connecticut. https://ncrge.uconn.edu/wp-content/uploads/sites/982/2016/01/NCRGE_EL_Lit-Review.pdf

National Association for Gifted Children (2011). *Glossary of terms*. www.nagc.org/resources-publications/resources/glossary-terms

National Center for Education Statistics (2013). *The total rate of poverty*. U.S. Department of Education. https://nces.ed.gov/pubs2013/2013037.pdf

National Center for Education Statistics (2017). Table 204.90. Percentage of public school students enrolled in gifted and talented programs by sex, race/ethnicity, and state: Selected years, 2004 through 2013–2014. *Digest of Education Statistics*. https://nces.ed.gov/programs/digest/d17/tables/dt17_204.90.asp

National Center for Education Statistics (2019). Table 204.10. Number and percentage of public school students eligible for free or reduced-price lunch, by state: Selected years, 2000–01 through 2018–19. *Digest of Education Statistics*. https://nces.ed.gov/programs/digest/d20/tables/dt20_204.10.asp

Nielsen, M. E. (2002). Gifted students with learning disabilities: Recommendations for identification and programming. *Exceptionality*, *10*(2), 93–111. https://doi.org/10.1207/S15327035EX1002_4

Olszewski-Kubilius, P., & Clarenbach, J. (2014). Closing the opportunity gap: Program factors contributing to academic success in culturally different youth. *Gifted Child Today*, *37*(2), 103–110. https://doi.org/10.1177/1076217514520630

Olszewski-Kubilius, P., & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly*, *62*(1), 37–55. https://doi.org/10.1177/0016986217738015

Olszewski-Kubilius, P. Steenbergen-Hu, S., Thomsen, D., & Rosen, R. (2017). Minority achievement gaps in STEM: Findings of a longitudinal study of Project Excite. *Gifted Child Quarterly*, *61*(1), 20–39. https://doi.org/10.1177/0016986216673449

Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, *60*(3), 159–171. https://doi.org/10.1177/0016986216643165

Pettit, S. K. (2011). Teachers' beliefs about English language learners in the mainstream classroom: A review of the literature. *International Multilingual Research Journal*, *5*(2), 123–147. https://doi.org/10.1080/19313152.2011.594357

Pettit, S. M. (2011). The effect of response to intervention and discrepancy testing in diagnosing English language learners with a reading learning disability. Master's Thesis, University of Texas.

Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.

Plucker, J. A., & Peters, S. J. (2018). Closing poverty-based excellence gaps: Conceptual measurement, and educational issues. *Gifted Child Quarterly*, *62*(1) 56–67. https://doi.org/10.1177/0016986217738566

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. https://cepa.stanford.edu/sites/default/files/reardon%20whither%20opportunity%20-%20chapter%205.pdf

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, *2*(3), 1–18. https://doi.org/10.1177/2332858416657343

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Sennett, R., & Cobb, J. (1972). *The hidden injuries of class*. Cambridge University Press.

Snellman, K., Silva, J. M., Frederick, C. B., & Putnam, R. D. (2015). The engagement gap: social mobility and extracurricular participation among American youth. *The ANNALS of the American Academy of Political and Social Science*, *657*(1), 194–207. https://doi.org/10.1177/0002716214548398

Stambaugh, T., & Wood, S. M. (2015). *Serving gifted students in rural settings*. Routledge.

Stormont, M., Stebbins, M. S., & Holliday, G. A. (2001). Characteristics and educational support needs of underrepresented gifted adolescents. *Psychology in the Schools*, *38*(5), 413–423. https://doi.org/10.1002/pits.1030

Trail, B. A. (2011). *Twice-exceptional gifted children*. Prufrock Press.

VanTassel-Baska, J., Feng, A., Swanson, J., Quek, C. & Chandler, K. (2009). Academic and affective profiles of low income, minority, and twice exceptional gifted learners: The role of gifted program membership in enhancing self. *Journal of Advanced Academics*, 20(4), 702–739. https://doi.org/10.1177/1932202X0902000406

Wai, J., Makel, M. C., & Gambrell, J. (2019). The role of elite education and inferred cognitive ability in eminent creative expertise: An historical analysis of the TIME 100. *Journal of Expertise*, *2*(2), 77–91. www.journalofexpertise.org/articles/volume2_issue2/JoE_2019_2_2_Wai.pdf

Weinfeld, R., Barnes-Robinson, L., Jeweler, S., & Shevitz, B. R. (2021). *Smart kids with learning difficulties: Overcoming obstacles and realizing potential* (2nd ed.). Routledge.

Wood, S. C. (2012). Examining parent and teacher perceptions of behaviors exhibited by gifted students referred for ADHD diagnosis using the Conners 3 (An exploratory study). *Roeper Review*, *34*(3), 194–204. https://doi.org/10.1080/02783193.2012.686426

Worrell, F. C. (2007). Ethnic identity, academic achievement, and global self-concept in four groups of academically talented adolescents. *Gifted Child Quarterly*, *51*(1), 23–38. https://doi.org/10.1177/0016986206296655

Worrell, F. C., & Dixson, D. D. (2018). Retaining and recruiting underrepresented gifted students. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children* (2nd ed., pp. 209–226). Springer.

Wyner, J. S., Bridgeland, J. M., & Dilulio Jr., J. J. (2009). *Achievement trap: How America is failing millions of high-achieving students from lower-income families*. Jack Kent Cooke Foundation and Civic Enterprises. www.jkcf.org/wp-content/uploads/2018/06/Achievement_Trap.pdf

Xiang, Y., Dahlin, M., Cronin, J., Theaker, R., & Durant, S. (2011) *Do high flyers maintain their altitude? Performance trends of top students*. Thomas B. Fordham Institute.

Conclusion to Part I on Assessments for Identification

Susan K. Johnsen and Joyce VanTassel-Baska

Introduction

The chapters contained in Part I focus on the types of assessments that might be used in identifying gifted students, particularly those from diverse groups. What types of tools are used in the identification process? What criteria should be used in selecting identification tools? How might these tools be used to identify students from underrepresented groups? How might these assessments be included in an equitable identification process that is effective in identifying students who need services and programming?

Summaries of Chapters

The following summaries of the five chapters provide research-based support for selecting and using qualitative and quantitative assessments in the identification process. Each of the chapters address different types of assessments, their advantages and disadvantages, and processes that need to be considered when planning and implementing an overall identification system. Throughout, emphasis is placed on approaches for identifying more students from underrepresented populations.

In the first chapter, Lindsay Lee and Scott Peters compared and contrasted universal screening, a two-phase process, with universal considerations, a one-phase process, where not only are data collected on all students but also all students are also considered. They argued that both of these approaches are more equitable than two-phase systems where parent and/or teacher referrals are required to enter the second phase. They provided criteria for selecting universal screeners such as alignment to the program, nomination validity, high reliability, quick administration, and inexpensive costs. They described a way for reducing costs by examining the assessments' reliability and the relationships between first phase and second phase instruments. They also described how lowering the phase-one cut score improves the sensitivity of the process and increases equity. They urged practitioners to consider the criteria presented in the chapter rather than

DOI: 10.4324/9781003285991-9

focusing on a particular instrument to use as a universal screener (e.g., nonverbal tests vs. achievement-tests). Lee and Peters concluded that improving sensitivity (i.e., the percentage of students the identification system correctly identifies) is key to improving equity.

Lakin et al. described a variety of quantitative assessments that might be considered in the identification process such as achievement and cognitive ability assessments. In the selection process, they emphasized that the assessment should match student needs and types of services. Using the Cattell-Horn-Carroll model of intelligence, they classified quantitative assessments on a continuum from those that assess broad abilities (e.g., general reasoning) to those that assess more school-acquired information (e.g., achievement). Practitioners, therefore, need to be familiar with the purpose of the test and how to interpret results. This requires understanding key measurement concepts such as norm- vs. criterion-referenced, local norms, common scores, measurement error, and measurement bias. In considering equity, they emphasized the importance of pre-identification activities such as creating Talent Pools, local norms, subgroup norms, cut-scores, and teacher misconceptions about the expectation for high academic achievement among gifted learners in all subjects. Their examples emphasized the importance of viewing quantitative data as a part of a body of evidence and interpreting the data holistically. They advised administrators to "be a talent scout not a deficit detector (p. 63)."

In the next chapter Little et al. focused on qualitative assessment tools. describing them as curriculum-based products and/or performance tasks. They emphasized how these tasks form a valuable and informative component of an overall identification system because they elicit performance that reflects the student's current developmental level and mastery within a content area or domain. To be useful, they described how these types of assessments need to link to gifted programming, be meaningful and engaging to the students, incorporate more advanced content and more complex steps, demonstrate a wide range of abilities and behaviors, and increase equitable access. However, they warned that using performance-based assessments may increase not only students from underserved populations but also students who are not from these groups. Similar to other assessments, administrators need to consider technical adequacy, costs, and need for additional professional preparation. Little et al., concluded by encouraging administrators to use performance tasks and products but to balance these authentic representations of students' classroom work with consistency of measures across classrooms and schools.

Ryser's chapter examined the usefulness of rating scales in identifying students who need services. The scales can be useful for providing information from multiple sources (e.g., teachers, parent, peers) in multiple contexts (e.g., school, home, after-school activities); however, they need to be based on actual observations of behavior, not vague descriptions. Training is critical to their reliability and validity. She suggested that bias can be minimized when educators consider three issues: low educator expectations, exclusive definitions, and lack of test fairness. When interpreting rating scales, Ryser emphasized the importance of using both quantitative and qualitative information from the scale so that the committee can access the specific types and examples of behaviors that are rated highly. Similar to the other authors, she described how universal screening, local norms, and psychometrically sound instruments improve the overall identification process and increase equity.

Equity is the focus of the final chapter in this section. VanTassel-Baska and Johnsen identified the characteristics of students from underrepresented groups (e.g., children of color, children from poverty, children whose heritage language is not English, children with disabilities, and other special populations). Each of these groups tend to be overlooked for specific reasons, which leads to their exclusion from services and programs. Moreover, students in these different subgroups are often dually labelled—poor and English language learners or twice exceptional with co-morbid conditions. When not served, these students experience achievement gaps—a widening of disparities across race and socioeconomic status as they progress in school. To examine some possible solutions, VanTassel-Baska and Johnsen interviewed coordinators of gifted programs in three states. They found that effective identification procedures generally included (a) tailoring identification and services to the strengths, needs, and interests of underrepresented groups; (b) using multiple approaches, (c) using data to improve approaches for future identification cycles, (d) broadening the focus to all underrepresented groups, and (e) emphasizing student outcomes for making changes.

Implementing a System of Identification

This conclusion to Part I synthesizes the best practices gleaned from the literature and effective practices identified by experienced school personnel that should assist other practitioners in reducing disparities in the identification of underrepresented students in school systems. The authors have assembled these processes in the context of overall program development so that educators and parents can see the interrelationship of identification to programming at the school level. The effort rests on several assumptions:

- That equal representation is possible, even given disparities in all phases of current identification systems;
- That technically sound instruments must be employed to improve the process;
- That the use of multiple approaches to solving the problem is better than using just one, so universal screening might be combined with the use of local norms or early programming at pre identification levels might be combined with curriculum-based measures;
- That the use of multiple sources (e.g., parents, teachers, peers) to complete inventories is superior to the judgment of just one teacher/educator; and
- That the use of current student performances and products adds a layer of authenticity to the process.

A complete identification system implies that there are a series of processes and steps that need to be enacted at both the planning and implementation stages. A central committee needs to be vested with the authority to plan and carry out all of the aspects of the process from selecting assessments to overseeing placement. Figure I.1 shows the nature of the identification cycle. It begins with the need to provide pre-identification opportunities for prospective students in the form of try-out curriculum that can be employed in K-2 classrooms. Then the district must select the instruments to be employed and how they will be administered, either in one or two phases that include universal screening. Other data need to be collected to frame profiles for each student to be considered for programs and decisions made on final selection. This stage of the process then is followed by placement decisions where the students are placed in a program well-matched to their aptitudes and interests. The teacher is informed of the

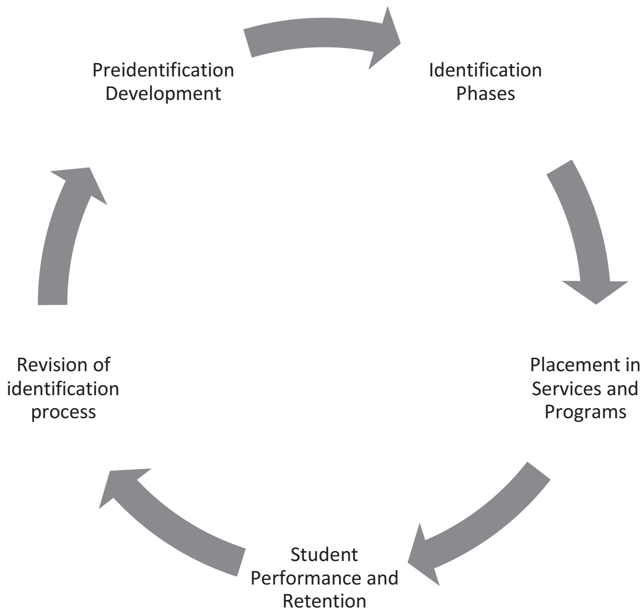

FIGURE I.1 The identification cycle for gifted programs and services

incoming student profiles and are provided commensurate professional learning opportunities to work with advanced learners. At the next stage in the cycle, the performance of students who were selected for the program are assessed and evaluated against the identified program outcomes. Retention decisions may need to be made also at this stage to determine if further differentiation needs to be made for student success, or if the student's rate of learning is not sufficient to benefit from the program to continue such as in a fast-paced math program. Finally, the identification process may be revised, based on needs identified during the cycle or on unanticipated results received from different data sources.

While the identification cycle in Figure I.1 shows the "big picture" of identification as an ongoing process, there are some specific considerations and steps to be taken within the cycle that require explication. Based on the authors of these chapters and the extant research, the steps for practitioners to consider are the following:

1. Select technically sound quantitative and qualitative screening instruments, based on the characteristics of the students and the services and programs. Identify how and when they will be administered.
2. Provide professional learning for all educators on the identification system and its processes.
3. Implement the identification system.
   - Create a pool that includes all students who have gone through universal screening on both quantitative and qualitative information depending on the domain (e.g., English/language arts, science, visual and performing arts).
   - Collect recommendations via a technically adequate checklist from teachers and parents on the abilities, interests, and psychosocial strengths who have been trained about the assessment's format, purpose, and key aspects of its scoring system.

- Collect quantitative information for each student (e.g., ability and achievement). Administer performance-based assessments, as appropriate, to each student in the pool who has demonstrated strength in a given area. Collect similar student products from each student or a portfolio of work to verify current levels of work.
- Establish local norms as needed for the assessments either at the school district or campus level.
- Establish a range of cutoff scores for each quantitative instrument that considers the standard error of measurement.
- Check for the issue of underrepresentation of target groups (i.e., race/ethnicity, poverty, EL, and twice exceptionality). Ensure that the range is representative of each subgroup in the district. Analyze the data by campus and use local norms to correct for problems of representation.

4. Establish a committee to organize the information for decision making and placement.
   - Create student profiles that include data from universal screening tests and tools, recommendations from teachers and parents, performance-based assessments, and work samples or portfolios.
   - Gather additional data by meeting with current and prospective teachers and review student profile data.
   - Create a profile of program options that match each student's profile.
   - The committee makes decisions related to placements, informs students and parents about placement decisions, oversees appeals and due process.
5. Assess the effectiveness of the system
   - Review the fidelity of implementation.
   - Analyze the data sources in respect to their appropriateness for purpose, their equity, and unintended consequences of their use.
   - Develop and administer a survey to educators on their perceptions of the identification system, especially its perceived strengths and weaknesses. Make changes as appropriate.
   - Examine the predictive validity of the instruments in terms of equity, student success in the program, and student retention.
   - Begin the next annual screening and identification process.

Concurrent Program Development with Identification

Identification processes, however, must be well-matched with program development activity. Appropriate program options need to be developed for the identified interests and strengths of the student population just identified. If program options are already in place, then they must be reviewed based on current student data. It is important to know, for example, if twice exceptional students will be a part of a given gifted cohort so that accommodations and modifications might be anticipated.

If pre-identification classes have not been developed, this area of service delivery should be a high priority. It might begin with the establishment of K-2 pilot classrooms that employ gifted curriculum in language arts, math, and science as options for students as a try-out mechanism for entering a pool of students for identification consideration. These classrooms might be set up with centers where gifted materials would be available for use. Teachers in these classrooms would need professional learning to prepare them for working with these students, using advanced and other differentiated materials. For example, reading materials should be included that are 2–4 grade levels advanced; math problem sets should also be graduated in degree of complexity beyond grade level expectations; opportunities to create science experiments should also be part of the science curriculum; and independent research options should be available in students' areas of interests. Moreover, enriched activities should be considered for students who are advanced in the arts.

Elementary programs that serve students with advanced abilities in reasoning, verbal, mathematical, and the arts areas of the curriculum should be established or expanded if only one or two of these areas has been developed. The choice of program grouping should be based on the issues within each district, but no program should operate without having gifted students working together in both general and specialized classrooms. Accelerated study also should be an integral part of the program delivery.

At the middle school level, programs should be advanced in each subject area where students have been assessed (e.g., verbal, math, the arts). Interdisciplinary options should be considered that address the needs and interests of targeted students (e.g., STEM, robotics, engineering). Opportunities for mentoring and participating in competitions such as History Day, science fairs, and writing contests should be explored.

At the high school level, options might include at least three Advanced Placement courses and/or the International Baccalaureate Program that could be offered at all levels of high school as students demonstrate readiness. Moreover, developing advanced courses that are prerequisites for these selective programs enhances the experience of advanced courses for underrepresented populations. It is also important to establish high school courses that address elective identified areas of interest, allowing students to be involved in independent research activities. For the promotion of career and college planning, schools might establish mentorships and internships across the curriculum spectrum for students in Grades 7–12.

Conclusion

As a field, gifted educators have always wanted to get the identification process right, finding the students who will benefit and succeed in advanced programs. Yet we know from many studies that we have been less successful than desired in this area of program development, often not identifying enough students who might have enjoyed success, especially among the underserved. We also have identified the wrong students, those who are compliant over those who speak and act out. And we have been woefully inadequate in providing optimal learning opportunities for students even when they have been identified. As the field advances in its knowledge of effective assessment measures and how to implement them effectively, may it also increase its use of equity in the process so that more learners may enjoy experiences tailored to their abilities and interests.

PART

II

Assessments for Learning Progress

Introduction to Part II on Assessments for Learning Progress

Joyce VanTassel-Baska and Susan K. Johnsen

This second section of the handbook on assessment focuses on how educators may gather information to examine students' learning progress. How do we know that gifted students have benefited from special services and programs? How do we know that there is an optimal match between these students' strengths and needs and the instruction they are receiving? What approaches to assessment have proven useful in answering these questions?

What is a Learning Progression?

The Model Core Teaching Standards (CCSSO & InTASC, 2011) define a learning progression as "increasingly sophisticated ways of thinking about a topic"…"support teachers' formative assessment practices and help teachers use learners' prior knowledge in productive ways" (p. 22). For purposes of this book, a learning progression may be defined as the pathway of advanced learning that gifted students are entitled to in public education and the processes and procedures necessary to document the differentiated learning they receive. When we consider what is meant by a "learning progression" for gifted learners, what components are important to address?

One of them is the framework that defines the goals and outcomes expected of these learners in schools. Because gifted learners are often typically functioning two grade levels beyond their age peers, they need to have curriculum experiences that optimally match their level of functioning. In order to judge the effectiveness of the learning at these advanced levels, the expectations for that learning must be defined upfront and matched to appropriate assessments.

Beyond the framework of stating goals and outcomes in a given domain, the outcomes need to be defined across the years that students are in school in each area of learning that the school is offering (i.e., cohesive). Thus, advanced learning outcomes must be defined from Pre-K-12 in all domains of study (e.g., language arts, mathematics, social studies, science, the arts). Other outcomes beyond content domains must also be

DOI: 10.4324/9781003285991-11

articulated to address social-emotional needs and college and career planning. Moreover, growth in psychosocial skills such as self-efficacy, metacognition, and resilience need to be addressed and monitored at critical stages in the schooling process (Subotnik et al., 2011).

Another component of a learning progression is the nature of the advanced learning provided, the types of opportunities that students may experience. In this aspect of the process, schools and districts must define a comprehensive set of services and program specializations they have instituted to meet the strengths and needs of gifted learners. Many districts combine classroom-based differentiation with special class instruction at particular levels, typically the secondary. In other districts, programs may include special classes or schools designated for the identified populations from early primary levels. In still others, it may be a combinational model of instruction that combines services in general education classroom settings with special options provided outside of the classroom. In still other districts, the use of acceleration in learning provides a continuous strand across all levels of learning, both in content and grade level.

Finally, a learning progression for any learner must be accountable to the state, to universities, and other entities that require evidence of student learning. The gifted student's advanced learning is no exception to this rule. So learning progressions must be marked by the use of advanced assessments commensurate with the nature, scope, and extent of learning that students accrue. The chapters in this segment of the book deliberately address such assessment tools and how they might be applied to document student growth at key stages of development.

The Bridge between Identification and Learning Progress

For directors of gifted programs, there is a need for having a process in place to transform the information from the identification process into a format that may be used by classroom teachers to develop curriculum and instructional strategies. While the end goal of testing and organizing a student profile is identification for gifted programs, the next step in the process is *placement* in gifted programs that match each student's interests, strengths and needs. Placement procedures should include the acknowledgement of assumptions held about gifted programs:

- That they are well-matched to students in respect to levels of aptitude, motivation, and interest;
- That they are sufficient in scope, depth, and complexity to respond to the needs of gifted students;
- That they offer advanced levels of instruction that are differentiated for these students; and
- That they are varied in order to be responsive to individual strength and needs, including those from underrepresented groups.

While the identification committee may feel the job is completed with selection, in reality, it must be followed by thoughtful decisions about placement. Moreover, it is the members of this committee who may have the best understanding of the students who have been selected for gifted programs and services. "Instructionally friendly" profiles developed for committee review are important data sources for teachers. Of special significance are: (a) the aptitude data that demonstrate the relative levels of student potential in critical areas of learning; (b) the student products and/or portfolio entries that show students' performance; and (c) the composite observations from educators about the particular interest, strengths, and needs of each student. These data need to be shared orally with relevant teachers in a context where they can ask questions and share their instructional goals for discussion. In this way, identification is the first important step toward learning progress processes that can put students on a productive path toward advanced learning in their areas of strength.

Purposes for Using Differentiated Assessment of Gifted Student Learning

However**,** differentiated assessment must accompany differentiated curriculum and instructional practices. For example, if a school district uses acceleration, each year groups of gifted learners would require advanced content assessments in each subject area in order to provide data on appropriate placement for the next year. In classrooms where advanced project work is the norm for identified students, assessments of projects should be applied to decisions about differentiation for these students in the classroom the following year. In an AP classroom, the required assessment for the course provides the evidence for credit and/or placement in college. The use of these various assessments is critical to ensure that a progression of learning for gifted students might occur in a seamless way across years without interruption or repetition.

Models of Advanced Learning

There are several models that school districts use to ensure that advanced learning is available across years for different groups of gifted learners. One model focuses on the major outcomes that districts try to develop in the gifted learner across Pre-K–12 (see Table II.1).

These outcomes may be addressed in existing content-based classrooms and/or be addressed in specialized settings beyond the classroom. These outcomes are long term in orientation, not occurring only in the expanse of a year but rather across multiple years or a lifetime of learning.

TABLE II.1 Cognitive and affective development in gifted programs

Cognitive outcomes	Affective outcomes
*Mastery-level work in areas of strength and interest (accelerated learning)	*Tolerance of self and others
*New areas of learning (i.e., novelty)	*Constructive use of humor
*Exploring interdisciplinarity	*Coping with being different
*Understanding human value systems	*Discriminating between the real and the ideal
*Discussions with intellectual peers	*Use of high-level sensitivity
*Applying complex levels of thought (i.e., higher level thinking)	*Developing relationships
*Creating divergent products	*Managing criticism
*Real-world problem solving	*Developing personal standards of excellence

Curriculum models represent another way to articulate learning progressions across the years of schooling. These curriculum models provide a sense of emphasis and focus on specific elements of learning, found to be important for gifted learners. Curriculum models provide an organizational structure for designing differentiated curriculum for the gifted. They represent what an ideal curriculum for the gifted might look like, based on a particular theory and design specifications. Some of the models have produced curricula Pre-K-12 that have been tested for effectiveness with gifted learners. As such, they remain useful tools to ensure that differentiation is integrated into the design of future curriculum products.

Models range from those that are content-based, such as the Integrated Curriculum Model (VanTassel-Baska, 1986), the Diagnostic Prescriptive Model (DP; Stanley, 1991), and the Parallel Curriculum Model (PCM; Tomlinson et al., 2002), to the Autonomous Learner Model (Betts & Neihart, 1986) that focuses on secondary programs that provide both cognitive and affective experiences outside of the core curriculum. Other models relate to conceptions of intelligence such as the Sternberg's Componential Model (Sternberg & Grigenko, 2003), which has been used to design units of study to match his theoretical conception of intelligence as analytical, synthetic, and practical. Studies of this model have focused on a comparison of gains of students exposed to different instructional approaches based on the model components. Some curriculum models focus on one aspect of curriculum design, such as using an instructional model of higher level thinking (e.g., Kaplan, 2009; Maker & Schiever, 2010) or problem-based learning (Gallagher, 2015) but do not provide a full design template (i.e., goals, outcomes, assessment).

Whether the model for learning is based on generic cognitive and affective outcomes or is based on a specific curriculum model that leads to establishing learner outcomes, school districts must have a consistent approach to providing advanced learning for gifted students. Only through such consistency and continuity can educators assess the rate and progress of gifted learners.

Research that Documents Advanced Learning of Gifted Students

For more than 100 years, research studies have used instruments to support accelerated learning and various grouping models as effective methods for serving gifted students in programs (see Steenbergen et al., 2016). Enrichment approaches that include differentiation applications in classrooms provide another level of evidence that demonstrates the positive learning outcomes of gifted learners in specific subject areas, in the use of project-based learning, and in affective areas of learning (Kim, 2016).

Subject-specific learning has been assessed in several studies over the last 20 years. Students have demonstrated significant growth gains in literary analysis and interpretation, persuasive writing, and linguistic competency in language arts in experimental gifted classes using the curriculum units based on the Integrated Curriculum Model in comparison to gifted groups not using such units (VanTassel-Baska et al., 2002). Findings from a 6-year longitudinal study examining the effects over time of using the William & Mary language arts units suggested that gifted student learning in grades 3–5 was enhanced at significant and educationally important levels in critical reading and persuasive writing. Repeated exposure over a 2–3-year period demonstrated increasing achievement patterns with the majority of stakeholders reporting the curriculum to be beneficial and effective (Feng et al., 2004). Other language arts studies produced results by using targeted pedagogical approaches to reading that enhanced fluency and comprehension for Title I students (Reis et al., 2008, 2011).

Studies on math curriculum have also shown positive results. According to the research, gifted students benefit most from math in alignment with the Common Core State Standards (CCSS). Specifically, these gains were most often found in higher level math problem solving and critical thinking when students were exposed to materials designed around the CCSS standards-based model (Gavin et al., 2007, 2009).

Gifted students in classrooms using Javits-funded science units outperformed gifted students in comparison classrooms on tests measuring their ability to apply the scientific method and demonstrate scientific reasoning skills (see Feng et al., 2004; VanTassel-Baska et al., 1998) and science concepts, content, and process (Kim et al., 2014). The studies also demonstrated the efficacy and motivational value of using a problem-based learning approach, embedded in an exemplary school science curriculum. More recent curricular interventions have focused on enhancing learning skills, using an alternative integrated model of instruction (e.g., Callahan et al., 2016) for students from low-income backgrounds.

A variety of formative and summative assessments are needed when appraising gifted students' learning and when differentiating the curriculum. Performance- and project-based assessments and tests of critical thinking are more likely to measure more complex thinking (Kim et al., 2014). Students are able to use written and creative products to reflect their ideas, academic identity, and intellectual understandings (Hall, 2007). Researchers have also found that gifted writers were able to rate student compositions similarly to experts in the field using the consensual assessment technique, which included review, feedback, and collaboration in the writing process (Kaufman et al., 2005).

Including students in the assessment process also improves the overall quality of student work. Newman (2004) found that when students were involved in self-assessment during the creative process their products were of higher quality. Likewise, Sriraman (2004) discovered that when mathematically gifted students were asked to reflect upon and analyze their own thinking processes, they were able to produce at a level characteristic of professional mathematicians. Being given the opportunity to design their own projects and assessments to measure achievement has also proven motivating to gifted learners (Thompson & McDonald, 2007). Moreover, the most creative and expressive products resulted from student-constructed assignments and assessments.

In the area of the assessment of social and emotional development, researchers have reported that for older gifted students, advanced coursework impacts their larger social lives. Contrary to popular belief, students in Advanced Placement or International Baccalaureate classes do not feel like they have to choose between academics or socializing (Foust et al., 2008); instead, they reported they had the best of both worlds. Students of color were also positively affected by relationships formed within the program (Walker & Pearsall, 2012). Students in these classes believed their peers motivated them to continue in rigorous coursework (Shiu et al., 2009). Specific interventions also appeared to enhance identity and resiliency. For example, Stutler (2011) used fictional literature to assist sixth-grade girls in dealing with adversity, and Whiting (2006) influenced the level of achievement among African American males using a scholar identity model.

Overview of Chapters

The authors in this section of the book have addressed the core questions in different ways, yet all describe assessments that document gifted students' differentiated and advanced progress and explore a better understanding of their learning outcomes. Some of the authors have addressed the use of alternative approaches to documenting learning such as through performance-based assessments, off-level assessments, and project-based assessments. Others have focused on the use of tools that are more sensitive to the abilities of these students as demonstrated in creative outcome measures and classroom-based curriculum tools. One chapter even focuses on the role of self-assessment in the process of learning for these students. Another chapter addresses the use of instruments that have gained credence in the larger education community to document advanced learning, specifically the Advanced Placement and International Baccalaureate examinations. Several chapters clarify why traditional standardized measures such as state assessment measures are insufficient to document the extent of learning that these students have mastered. Another of the chapters highlights the kind of accommodations needed to document the advanced learning of gifted students with disabilities, providing ideas for tailored assessment protocols that may provide more appropriate ways to document the growth of these students. Finally, one of the chapters addresses the long-term issues of demonstrating learning for the gifted that leads to competence in domains and eminence in the larger society.

Conclusion

The importance of assessing the learning of gifted students both continuously and consistently cannot be overstated. The Common Core Standards of Learning set up the basis for assessments in the majority of school districts nationally. Yet the accompanying state assessments often underestimate the scope, the type, and the level of learning of which gifted students are capable. Utilizing alternative assessment models and approaches provides an excellent way to gain a deeper appreciation for the true learning that gifted programs may be producing. The chapters that follow provide important ideas for doing just that.

References

Betts, G. T., and Neihart, M. (1986). Implementing self-directed learning models for the gifted and talented. *Gifted Child Quarterly*, *30*(4), 174–177. https://doi.org/10.1177/001698628603000406

Callahan, C. M., Moon, T. R., Oh, S., Azano, A. P., and Hailey, E. P. (2016). What works in gifted education: Documenting the effects of an integrated curriculum/instructional model, *American Education Research Journal*, *52*(1), 137–167. https://doi.org/10.3102/0002831214549448

Council of Chief State School Officers and Interstate Teacher Assessment and Support Consortium (2011). *InTASC model core teaching standards: A resource for state dialogue*. https://ccsso.org/sites/default/files/2017-11/InTASC_Model_Core_Teaching_Standards_2011.pdf

Feng, A., VanTassel-Baska, J., Quek, C., Bai, W., & O'Neill, B. (2004). A longitudinal assessment of gifted students' learning using the integrated curriculum model: Impacts and perceptions of the William and Mary language arts and science curriculum. *Roeper Review*, 27(2), 78–83. https://doi.org/10.1080/02783190509554294

Foust, R. C., Hertberg-Davis, H., & Callahan, C. M. (2008). "Having it all" at sleep's expense: The forced choice of participants in Advanced Placement courses and International Baccalaureate programs. *Roeper Review*, *30*(2), 121–129. https://doi.org/10.1080/02783190801955293

Gallagher, S. A. (2015). Adapting problem-based learning for gifted students. In F. A. Karnes & S. M. Bean (Eds.), *Methods and materials for teaching the gifted* (4th ed., pp. 413–444). Prufrock Press.

Gavin, M. K., Casa, T., Adelson, J. L., Carroll, S. R., & Sheffield, L. J. (2009). The impact of advanced curriculum on the achievement of mathematically promising elementary students. *Gifted Child Quarterly*, *53*(3), 188–202. https://doi.org/10.1177/0016986209334964

Gavin, M. K., Casa, T. M., Adelson, J. L., Carroll, S. R., Sheffield, L. J., & Spinelli, A. M. (2007). Project M^3: Mentoring mathematical minds—A research-based curriculum for talented elementary students. *Journal of Advanced Academics*, *18*(4), 566–585. https://doi.org/10.4219/jaa-2007-552

Hall, H. R. (2007). Poetic expressions: Students of color express resiliency through metaphors and similes. *Journal of Advanced Academics*, *8*(2), 216–244. https://doi.org/10.4219/jaa-2007-355

Kaplan, S. N. (2009). The grid: A model to construct differentiated curriculum for the gifted. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted and talented* (2nd ed., pp. 235–252). Prufrock Press.

Kaufman, J. C., Gentile, C. A., & Baer, J. (2005). Do gifted student writers and creative writing experts rate creativity the same way? *Gifted Child Quarterly*, *49*(3), 260–265. https://doi.org/10.1177/001698620504900307

Kim, K. H., VanTassel-Baska, J., Bracken, B. A., Feng, A., & Stambaugh, T. (2014). Assessing science reasoning and conceptual understanding in the primary grades using standardized and performance-based assessments. *Journal of Advanced Academics*, *25*(1), 47–66. https://doi.org/10.1177/1932202X14520946

Kim, K., VanTassel-Baska, J., Bracken, B. A., Feng, A., Stambaugh, T., & Bland, L. (2014). Project Clarion: Three years of science instruction in Title I schools among k-third grade students. *Research in Science Education*, *42*(5), 813–829. https://eric.ed.gov/?id=EJ978037

Kim, M. (2016). A meta-analysis of the effects of enrichment programs on gifted students. *Gifted Child Quarterly*, *60*(2), 102–116. https://doi.org/10.1177/0016986216630607

Maker, C. J., & Schiever, S. W. (2010). *Curriculum development and teaching strategies for gifted learner*s (3rd ed.) PRO-ED.

Newman, J. L. (2004). Talents and type IIIs: The effects of the talents unlimited model on creative productivity in gifted youngsters. *Roeper Review*, *27*(2), 84–90. https://doi.org/10.1080/02783190509554295

Reis, S. M., Eckert, R. D., McCoach, D. B., Jacobs, J. K., & Coyne, M. (2008). Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. *Journal of Educational Research*, *101*(5), 299–315. https://doi.org/10.2300/JOER.101.5.299-315

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, *48*(2), 462–501. https://doi.org/10.3102/0002831210382891

Shiu, A., Kettler, T., & Johnsen, S. K. (2009). Social effects of Hispanic students enrolled in an AP class in middle school. *Journal of Advanced Academics*, *21*(1), 58–82. https://doi.org/10.1177/1932202X0902100104

Sriraman, B. (2004). Gifted ninth graders' notions of proof: Investigating parallels in approaches of mathematically gifted students and professional mathematicians. *Journal for the Education of the Gifted*, *27*(4), 267–292. https://doi.org/10.4219/jeg-2004-317

Stanley, J. C. (1991). An academic model for educating the mathematically talented. *Gifted Child Quarterly*, *35*(1), 36–42. https://doi.org/10.1177/001698629103500105

Steenbergen-Hu, S., Makel, M., C., & Olszewski-Kubilius, P. (2016). What 100 years of research says about the effects of ability grouping and acceleration on K-12 students' academic achievement: Findings of two second order meta-analyses. *Review of Educational Research*, *86*(4), 849–899. www.jstor.org/stable/44668238

Sternberg, R. J., & Grigenko, E. L. (2003). Teaching for successful intelligence: Principles, procedures, and practice. *Journal for the Education of the Gifted*, *27*(2–3), 207–228. https://doi.org/10.1177/016235320302700206

Stutler, S. L. (2011). Gifted girls' passion for fiction: The quest for meaning, growth, and self-actualization. *Gifted Child Quarterly*, *55*(1), 18–38. https://doi.org/10.1177/0016986210383979

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Thompson, D. D., & McDonald, D. M. (2007). Examining the influence of teacher-constructed and student-constructed assignments on the achievement patterns of gifted and advanced sixth-grade students. *Journal for the Education of the Gifted*, *31*(2), 198–226. https://doi.org/10.4219/jeg-2007-676

Tomlinson, C. A., Kaplan, S. N., Renzulli, J. S., Purcell, J., Leppien, J., & Burns, D. (2002). *The parallel curriculum: A design to develop high potential and challenge high ability learners*. Corwin Press.

VanTassel-Baska, J. (1986). Effective curriculum and instructional models for talented students. *Gifted Child Quarterly*, *30*(4), 164–169. https://doi.org/10.1177/001698628603000404

VanTassel-Baska, J., Bass, G., Reis, R., Poland, D., & Avery, L. (1998) A national pilot study of science curriculum effectiveness for high ability students. *Gifted Child Quarterly, 42*(4)*,* 200–211. https://doi.org/10.1177/001698629804200404

VanTassel-Baska, J., Zuo, L., Avery, L. D., & Little, C. A. (2002). A curriculum study of gifted student learning in the language arts. *Gifted Child Quarterly*, *46(*1), 30–44. https://doi.org/10.1177/001698620204600104

Walker, S. A., & Pearsall, L. D. (2012). Barriers to Advanced Placement for Latino students at the high-school level. *Roeper Review*, *34*(1), 12–25. https://doi.org/10.1080/02783193.2012.627549

Whiting, G. W. (2006). From at risk to at promise: Developing scholar identities among Black males. *Journal of Advanced Academics*, *17*(4), 222–229. https://doi.org/10.4219/jsge-2006-407

CHAPTER

7

The Assessment of Domain-Specific Creativity

Person, Process, and Product

Todd Kettler and Kristen N. Lamb

Assessment is a complex process inherent in all educational systems including gifted education. A fundamental purpose of learning assessments is to ensure quality education (Archer, 2017; Black & Wiliam, 1988). In other words, assessment tools and practices are essential to documenting quality, needs, and progress in students and in programs. Assessment maintains a prominent role in the National Association for Gifted Children (NAGC) Programming Standards (2019). Those standards articulate three general purposes for assessment: (a) identification of students to participate in gifted and talented education, (b) documentation of students' learning progress within gifted education programs, and (c) evaluation of programming and services. In this chapter, we are applying the second purpose, documentation of learning, to the narrow performance space of domain-specific creativity. The purpose of this examination is to provide an assessment framework and applicable tools to those seeking to measure students' growth in creative thinking within gifted education programs and talent development trajectories.

Historically, educators have categorized assessments to document learning into two general forms—formative and summative assessment (Brookhart, 2001). However, some contend that those categorizations are insufficient to adequately guide assessment practices (Newton, 2007; Schildkamp & Kuiper, 2010). A single assessment can be both formative and summative. The terms relate to categories of purpose rather than classification of the assessment tools and processes and are therefore more nuanced. Newton (2007) articulated 18 purposes for educational assessment that capture the nuance of purposes and users of assessments more thoroughly than the basic formative and summative distinctions.

Applying Newton's (2007) framework of purpose-driven educational assessment, we suggest the following categories or purposes of domain-specific creativity assessments. These four uses of domain-specific creativity together support quality gifted education programs and services. Assess domain-specific creativity:

- To identify student (or group) learning needs and to guide subsequent teaching and learning. (*Formative Assessment*)

DOI: 10.4324/9781003285991-12

- To determine whether students are making sufficient progress or achievement over time—attainment of learning goals in gifted education. (*Student Monitoring*)
- To support student talent development and guide future decisions about subsequent education and career pathways. (*Guidance*)
- To evaluate the success of the gifted education curriculum and instruction focused on developing domain-specific creativity skills. (*Program evaluation*)

The Important Role of Creativity in Education

Rapid economic transformation has elevated creativity to an essential skill. This transformation has ushered society from the industrial economy to an economy that thrives from intelligent and creative input. Florida (2006) refers to this emerging economy as the knowledge economy, or creative economy. Presently, about nine out of ten college graduates in America work within the creative sector, and these sectors account for more than half of the workforce in leading metropolitan areas (Florida, 2019). Additionally, approximately 7.6 million people in the United States work within the creative sector; that is about 4% of total employment and 15% of monthly income in the U.S. (Florida & Seman, 2020).

As the demand for creative thinking and creative talent has increased in priority for many nations across the globe, many individuals leave schools extremely unprepared for employment in the knowledge economy that requires creative thought. To address the increasing need for creative thinkers, multiple stakeholders have called for educational reform that includes creativity as a central goal of education (Florida, 2006; Lamb, 2020a; P21, n.d.; Robinson, 2006). One way to address the need for creative thinkers is through the intentional development of creative talent.

Overall, creative talent is expressed within two dimensions: creative performance (e.g., theatre, visual art, music, etc.) and creative production (e.g., medicine, technology, physics, etc.). These dimensions provide information related to how creative talent is developed and expressed (Subotnik et al., 2011). Though a student's general abilities play an important role in talent development, it is their abilities in specific domains that matter more. Because creative talent manifests in domain-specific ways (Piirto, 2022), the identification and development of creative talent also varies by discipline.

Many states account for creativity as a component of giftedness in their state definitions (e.g., Alabama, Texas) and provide guidelines for school districts to use general creativity assessments to identify students for participation in gifted and talented education programs. However, even when students are identified, based on their creativity, they are often not served in domain-specific areas of creativity. In other words, students may be identified for their creative potential but rarely are placed in programs that support the development of that creative talent within a particular discipline. Yet opportunities for creative talent development are crucial to a student's trajectory. This chapter aims to address this issue through a discussion of how to systematically assess student learning progress in domain-specific creativity, consistent with talent development principles.

Understanding Domain-Specific Creativity

Creativity is a complex construct that has evolved over time. Moreover, the concept of creativity has also been romanticized and misunderstood, yielding myths and misconceptions of what it means to think or behave in creative ways. Some evidence suggests that teachers often have misperceptions of the meaning of creativity (Bereczki & Kárpáti, 2018; Mullet et al., 2016). Teachers tend to believe creativity is a stable trait that cannot be changed through training and practice. They also tend to associate creativity with the arts or design but not core curriculum (Andiliou & Murphy, 2010).

Defining Creativity for Education

Like many complex constructs, creativity has a plethora of definitions. Using meta-synthesis techniques, Plucker et al. (2004) defined creativity as, "the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" (p. 90). Two facets of the definition that yield particular relevance to education are (a) the requirement for a perceptible product that becomes the object of creative assessment and (b) the social context which implies domain-specific learning and performance spaces. With a somewhat different emphasis, Mumford et al. (2018) defined thinking and behaving in a creative problem-solving context, "Creativity involves the production of original, high quality, and elegant solutions to a certain class of problems—novel, complex, and ill-defined, or poorly structured problems" (p. 147). The educational relevance of the Mumford et al. (2018) definition is that certain types of problem solving tend to produce creative responses, and those problems require learning designs that ask students to respond to ill-defined or poorly structured problems within domain-specific learning and performance spaces. Both definitions of creativity reflect cognitive creativity, manifested in domain-specific products or performances. The first step in assessment of domain-specific creativity is clearly defining creative thinking and behaving and understanding how that would be reflected in a domain-specific product or performance.

Educational creativity is domain-specific; thus, assessment of student performances and growth must be conceived in a domain-specific framework. Amabile (1982, 1996) developed the componential model of creativity comprised of three components: (a) domain knowledge and skills, (b) creativity-relevant processes, and (c) task motivation. More recently, Amabile and Pratt (2016) updated that model as a dynamic componential model of creativity and innovation to include a fourth element, (d) the social environment in which the person is working. Thus, educational creativity results when a student combines the knowledge and skills of a domain (mathematics, science, language, social studies) with creative cognition, and the motivation to respond in novel ways in a learning context. The educational model of domain-specific creativity posits that as students respond to open-ended tasks or ill-defined problems, their creative response reflects varying degrees of each of the three components (cognitive creativity skills, domain-specific knowledge and skills, and motivation or willingness to respond in novel yet appropriate ways to the task). Therefore, assessment of domain-specific creativity must involve all three areas. Educators might presume that exceptional evidence of domain-specific creativity indicates high levels of all three components. Yet, less than exceptional creative products could result from varying levels of any or all three. Knowing these assumptions allows educators to merge curriculum, instruction, and assessment with the intent of teaching and measuring domain-specific creative achievement.

Creativity in the Core Curriculum

When the scientific study of creativity was formalized in the middle of the twentieth century (Guilford, 1950), creativity was largely considered a domain-general phenomenon.

Researchers studied creative people and their traits (person), creative processes such as creative thinking or creative problem solving (process), and creative products, the artifacts of creative people using creative processes (product) (Lamb, 2020b; Rhodes, 1961; Said-Metwaly et al., 2017). By the end of the twentieth century, the assumption of domain-generality was widely questioned as data across multiple studies pointed toward domain-specificity (Baer, 1998; Han 2003; Piirto, 2000, 2014). Additionally, Avitia and Plucker (2014) conducted a content analysis of creativity textbooks and found that almost every textbook described the domain-specific nature of creativity. As theory and research in creativity have converged and affirmed the domain-specific nature of creativity, the assessment of creativity must also be domain-specific (Baer, 2015).

Assessing Domain-Specific Creativity: Person

There are three ways of measuring domain-specific creativity consistent with creativity research traditions: creative person, creative process, and creative products. We describe three ways to assess domain-specific creativity at the person level: (a) creative strengths profile, (b) rating scale for assessing student creative traits, and (c) students' creative self-beliefs. The primary purpose for assessing domain-specific creativity at the person level is guidance—assessment to support talent development and guide future decisions about subsequent education and career pathways. A second purpose for assessing domain-specific creativity at the person level is student monitoring. In what ways is the student showing progress over time on these person-based measures?

Creative Strengths Profile

Treffinger et al. (2013) developed the Creative Strengths Profile (CSP) to measure ways that students demonstrate creative characteristics and strengths on educational tasks. Treffinger and colleagues described seven characteristics of the assessment tool: flexible, developmental/dynamic, focused on strengths, diagnostic, functional, multiple sources of data, and action-oriented. It should be used within domains, and it should provide guidance on subsequent programs or services. It can also be used over time to monitor student growth. Because the CSP is flexible, teacher teams or district gifted education teams can determine which data are collected and tracked on the profile. The goal of the CSP is to collect accurate and useful information over time. Strength-based assessments have been practiced for years with most applications in special education (Van Den Berg & Grealish, 1996). Students and families generally find strength-based approaches to skill development to be a positive experience (Epstein & Sharma, 1998).

Rating Scale for Assessing Student's Creative Traits

The Rating Scale for Assessing Student's Creative Traits (Proctor & Burnett, 2004; see Table 7.1) helps teachers measure creative traits in the students they observe in class. The scale measures nine traits that all load on a single factor of creativity. Proctor and Burnett (2004) found the instrument to have a high level of reliability (α = .93). The scale has demonstrated consistency over time with test-retest correlation coefficients estimated at r = .66. Kettler and Bower (2017) adapted and used the Proctor and Burnett rating scale in an elementary language arts program and also found construct validity for the scale. Teachers' ratings using the scale were related to students' creative writing performances

TABLE 7.1 Rating scale for assessing student's creative traits

Descriptor	Indictors	Rarely	Sometimes	Often
Fluent Thinker	The student is full of ideas; finds different ways of doing things; answers questions fluently and readily; hypothesizes easily, generally possesses high verbal fluency; can list, tell/retell, label and compile easily; answers (fluently) questions such as how many? Why? What are the possible reasons for? Just suppose…?	**1**	**2**	**3**
Flexible Thinker	The student can solve, change, adapt, modify, magnify, rearrange, reverse, and improve; is versatile and can cope with several ideas at once; is constructive and mentally builds and rebuilds; is sensitive to new ideas and flexible in approach to problems; can tolerate ambiguity.	**1**	**2**	**3**
Original Thinker	The student can create, invent, make up, construct, substitute, combine, compose, improve, and design; is attracted by novelty, complexity, mystery; ask What if? questions.	**1**	**2**	**3**
Elaborate Thinker	The student can enlarge, extend, exchange, replace and modify; goes beyond assigned tasks; sees new possibilities in the familiar; embellishes stories/situations.	**1**	**2**	**3**
Intrinsically Motivated Student	The student often seeks knowledge independently; does a job well for its own sake, not for rewards; appears to enjoy learning for learning's sake.	**1**	**2**	**3**
Curious Student Immersed in the Task	The student tries to discover the unusual or find out more about a topic of interest; unable to rest until the work is complete; possesses a sense of wonder and intrigue; possess a high energy level; is adventurous and engages in spontaneous action; can uncover, investigate, question, research, analyze, see out and ponder.	**1**	**2**	**3**
Risk Taker	The student will challenge, criticize, judge, question, dispute, and decide; not afraid to try new things; not afraid to fail; can rank and give reasons, justify, defend, contrast and compare, devise a plan, make a choice between.	**1**	**2**	**3**
Imaginative or Intuitive Thinker	The student will fantasize, create, compose, invent, suppose, dramatize, design, dream, wish; is perceptive and sees relationships; can make mental leaps from one idea to another and from the known to the unknown.	**1**	**2**	**3**
Engages in Complex Task and Enjoys a Challenge	The student can evaluate, generalize, abstract, reflect upon, move from concreate to abstract, move from general to specific, converge, and has problem tolerance; is not easily stressed; does not give up easily; often irritated by the routine and obvious.	**1**	**2**	**3**

For each descriptor, rate the student on the nine creativity traits based on performance in your [math / language arts / science / social studies] class.

Rarely = less than 30% of the time
Sometimes = 30–70% of the time
Often = more than 70% of the time.

Adapted from Kettler & Bower (2017) and Proctor & Burnett (2004)

(r = .27) and the scale also showed a strong positive correlation with the Renzulli Scales– Creativity (Renzulli et al., 2013; r = .83) and the Scales for Identifying Gifted Students– Creativity (SIGS; Ryser & McConnell, 2004; r = .80). Interrater reliability indicated that the scale can be consistently interpreted by teachers with agreement rates ranging from 55% to 78% across the nine rating items.

School systems could use the Rating Scale for Assessing Student's Creative Traits in any domain, including arts and career and technology courses. The assessment is based on what the teacher observes in domain-specific environments. Students' scores yield a profile across each of the nine traits that can serve all four of the above stated assessment purposes: (a) formative assessment, (b) student monitoring, (c) guidance, and (d) program evaluation. Teachers' use of the rating scale is straightforward and their ratings may become more consistent with basic training involving definitions of creativity and the domain-specific creative behaviors. Systematic use of the rating scales could include all teachers at a school or in a gifted program completing the scales during creativity assessment windows two or three times per year in order to obtain longitudinal data and monitor student growth in creative performance across domains.

Student Self-Beliefs about Creativity

Creative self-efficacy is a student's belief related to the ability to respond in creative ways to academic work. Creative self-beliefs are conceptual and empirical precursors to creative behaviors (Karwowski et al., 2019), and self-belief constructs are generally considered malleable through both internal and external influences such as educational interventions (Karwowski & Beghetto, 2018; Karwowski et al., 2015). Students with higher levels of creative self-efficacy take more intellectual risks, seek opportunities for creative expression, and expend more effort on creative tasks (Tierney & Farmer, 2002). Beghetto (2006) developed a short creative self-efficacy assessment and found it reliable (α = .86) with adolescent students. The assessment includes three items (a) "I am good at coming up with new ideas," (b) "I have lots of good ideas," and (c) "I have a good imagination" (Beghetto, 2006, p. 450), which are scored on a 5-point Likert scale (1-not true, 2-mostly not true, 3-somewhat true, 4-true, 5-very true). The validity of creative self-efficacy (including the Beghetto scale) has been supported through meta-analysis (Haase et al., 2018). Sixty effect sizes from 41 papers (N = 17,226) indicated a medium effect size (r = .39) between creative self-efficacy and creativity measures. The relationship was slightly stronger when creative self-efficacy was compared to measures focusing on the creative person (r = .47; Haase et al., 2018). Students scoring higher on the creative self-efficacy scale also held more positive beliefs about their academic abilities, participated in more after-school academic activities, and were more likely to indicate that they planned to attend college. Beghetto's scale has been used in school-based research (e.g., Kaufman, 2019; Rubenstein et al., 2018) with students and can easily be used as part of a battery of ongoing assessments of domain-specific creativity (Beghetto & Baxter, 2012).

Creative personal identity measures to what degree creativity is important to a student's self-description, and creative personal identity is considered a component (along with creative self-efficacy) of a student's creative self-concept (Karwowski, 2013). Karwowski (2013) developed a five-item scale to measure creative personal identity on a 5-point Likert scale (1-definitely not, 2-not, 3-somewhat, 4-yes, 5-definitely yes). The items include (a) "I think I am a creative person;" (b) "My creativity is important for who I am;" (c) "Being a creative person is important to me;" (d) "Creativity is an important part of me;" and (e) "Ingenuity is a characteristic which is important to me" (Karwowski, 2013, p. 65). Karwowski found the scale reliable with adolescents and adults (α = .85). Creative personal identity is positively related to creative achievement, and creative personal identity has been found to be generally more malleable than personality traits (Karwowski, 2013). Thus, when students encounter positive experiences with domain-specific creativity, it is possible to increase their creative personal identity in that domain over time.

Measuring Domain-Specific Creativity: Process

Creative processes are cognitive functions associated with thinking creatively. For instance, students use creative processes when they think divergently, problem solve, or perform in original ways. Meta-analytic research (Scott et al., 2004), examining 70 empirical studies on the effectiveness of creativity training, found that creative processes can be improved in students and adults through specific training and practice. Overall, the study indicated a medium-to-large effect (d = .68) for such creativity training. Specifically, looking at three general aspects of creativity, they found a range of positive effects of training for divergent thinking (d = .75), problem solving (d = .84), and creative performance (d = .35). The most effective creativity training approaches emphasize the use of specific cognitive processes and provide students with explicit guidelines, checklists, and feedback. The length of time of the creativity training and the amount of practice opportunities were also positively correlated with increased creative outcomes (Scott et al., 2004). Most research on creativity training has been domain-general. Implementing these teaching and assessment practices in domain-specific contexts requires teachers to identify specific cognitive processes and demonstrate how those processes are used in domain-specific contexts. (Beghetto & Kaufman, 2010; Grigorenko et al., 2008). The taxonomy of creative thinking and inquiry learning through creative problem solving are two ways to teach and assess domain-specific creative processes.

Taxonomy of Creative Thinking

To make more explicit the cognitive processes of creative thinking, Kettler et al. (2018) developed the taxonomy of creative thinking (see Table 7.2). This taxonomy identifies 16 specific creative thinking skills across five general cognitive areas. The taxonomy is effective for integrating specific creative thinking processes with domain-specific content. For instance, a fifth grade science standard (Texas) requires that "students identify alternative energy resources such as wind, solar, hydroelectric, geothermal, and biofuels." When a teacher designs instruction to teach this standard, they would integrate one or more of the creative thinking processes into the unit. For example, the teacher may include 4.1, "students will recognize and describe problems that could be solved through the use of alternative energy resources."

In the most effective scenario, school systems would integrate the taxonomy of creative thinking with the core curriculum domains across all grade levels. Domain-specific teams could deconstruct each of the 16 processes and even further articulate what each of the creative thinking skills would look like in mathematics, science, language arts, and social studies. Each of the skills could be systematically integrated into the domain-specific curricula, and professional learning opportunities would support those teachers' abilities to explain, model, and assess each of the creativity processes. By specifying the

TABLE 7.2 Taxonomy of creative thinking

1. Idea Generation	1.1: Students will generate ideas that reflect original thinking about the content of study.
	1.2: Students will effectively use a wide range of creation techniques.
	1.3: Students will clearly communicate the ideas that they develop.
	1.4: Students will develop alternative explanations for events or phenomena within the content of study.
	1.5: Students will construct theories to explain phenomena within the content of study.
2. Idea Elaboration	2.1: Students will elaborate their own ideas as well as the ideas of others by adding more details.
	2.2: Students will analyze and define their own ideas as well as the ideas of others to make ideas more accurate or effective.
	2.3: Students will analyze and evaluate alternative explanations or alternative points of view.
3. Idea Connections	3.1: Students will make connections between new ideas and existing ideas as well as between multiple existing ideas.
	3.2: Students will combine parts of existing idea to generate original extensions of those ideas.
4. Problem Solving	4.1: Students will recognize and describe problems that could be solved.
	4.2: Students will apply problem-solving protocols to generate creative solutions to problems.
	4.3: Students will predict outcomes in hypothetical models.
5. Original Work	5.1: Students will produce products that reflect originality and are authentic to the domain of study.
	5.2: Students will use technology to generate innovative outcomes within the domain of study.
	5.3: Students will demonstrate originality, imagination, and innovation thinking in their work within the domain of study.

From Kettler, T., Lamb, K. N., & Mullet, D. R. (2018). *Developing creativity in the classroom: Learning and innovation for 21st century schools*. Prufrock Press. Used with permission.

creative thinking processes that would be integrated into teaching and learning, teachers would assess domain-specific creativity across each of those 16 specific processes as they are applied to domain-specific content. In this way, domain-specific assessment of creativity is seamlessly achieved through curriculum-based assessments (Grigorenko et al., 2008). Curriculum-based assessments on these 16 creative processes serve three specific assessment purposes. Formative assessment could provide teachers and students with data to inform interventions, focused instruction, or differentiation. Teachers and administrators could monitor students' growth in each of the five areas and 16 creative processes over time. Lastly, program administrators could use the curriculum-based assessments of creative processes as a component of gifted education program evaluation.

Creative Problem Solving

Creative problem solving refers to the ability to solve a certain class of problems that is ill-structured or vaguely defined and requires solutions that are elegant and novel. Creative problem solving is regarded as one of the most desired practical skills in the complex world of knowledge-economy work. There are several cognitive processes associated with creative problem solving. Domain-specific knowledge provides a conceptual framework for creative problem solving, and students with well-organized conceptual knowledge structures typically produce the most creative solutions to ill-structured problems (Mumford et al., 2012).

One way to assess creative problem solving is to assess the outcome, the product of the problem-solving process. However, teachers can also assess students' engagement in the creative problem-solving process. The process involves applying domain-specific knowledge through divergent thinking, convergent thinking, and conceptual combination. When students engage in domain-specific divergent thinking, their application of the content knowledge may reflect originality, elaboration, fluency, and flexibility. Their convergent thinking should involve the use of tools to evaluate and narrow ideas and concepts to determine the most viable solutions or application of the domain-specific knowledge and skills. Students should demonstrate their understanding of the concepts of the domain in both the problem definition and possible solutions. The following questions should guide how teachers assess the students' creative thinking process during creative problem solving:

- How thoroughly did the student elaborate the problem through the application of domain-specific knowledge and skills?
- In what ways did the student apply flexible thinking to explore the problem from a unique or novel perspective?
- In what ways did the student apply domain-specific knowledge and skills in the generation of novel or original ideas as possible ways to address the problem?
- How many possible solutions did the student generate before engaging in solution evaluation?
- How thoroughly did the student consider elements of the solution implementation to demonstrate novel and flexible application of domain-specific knowledge and skills?

Assessing domain-specific creativity through cognitive processes requires focusing on the process in addition to the outcome. Specifically defining creative thinking processes and generating examples of how they are applied in domain-specific ways prepares teachers to monitor students' progress in the development of creative thinking and performance in each domain. Systematically defining creative processes, integrating them into the curriculum, and assessing those processes with curriculum-based assessment should be thoroughly supported through ongoing professional learning within domain-specific teams.

Measuring Domain-Specific Creativity: Product

Perhaps most emphasized in school settings, students' products and performances are assessed to measure and document growth and mastery. Assessing creative products should consider the domain in which the product is associated. The primary purpose for assessing domain-specific creativity at the product level is domain expertise—assessments support talent development toward expertise through authentic artifacts of learning. A second purpose for assessing domain-specific creativity at the product level is authentic learning—assessments support talent development through authentic learning experiences. In this area, the student demonstrates progress over time on product-based measures such as the consensual assessment technique and domain-specific creativity rubrics.

The Consensual Assessment Technique

The Consensual Assessment Technique (CAT) is highly reliable and one of the most widely used instruments to measure domain-specific creative products (Amabile, 1982; Kettler et al., 2018). The CAT has demonstrated high "interjudge reliability" for a variety of creative tasks across multiple domains (e.g., psychologists (α = .85), poets (α = .88), math students (.87), teachers (α = .87) (Hennessey, 1994, p. 195). Moreover, the CAT has documented use in school settings with promising results (i.e., stories (α = .94), personal narratives (α = .96), and poems (α = .87); Baer et al., 2004). Under this technique, students are tasked with creating a domain-specific product. The CAT then relies on experts within a discipline to judge the quality of a creative product: "A product is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created" (Amabile, 1982, p. 1000). It is the versatile nature of the CAT that leads to its applicability in various domains.

Procedures for the CAT mimic procedures used to evaluate creative products in real world settings (Said-Metwaly et al., 2017). Incorporating the CAT as an assessment tool for creative products supports domain-specific creative talent development by providing students with real world learning experiences. Under this approach, students can receive expert feedback that mimics the nature of the discipline. The procedures for the CAT include (adapted from Kettler et al., 2018):

- Select judges based on their experience in a particular domain (the experience level does not have to be the same across all judges selected).
- Creative products should be presented to judges in a randomized order.
- Instruct judges to independently evaluate creative products (no training or objective criteria should be provided).
- Ask judges to also assess other dimensions of the creative product (e.g., aesthetics, cost).
- Collect the completed evaluations and compare the judges' scores to determine interrater reliability (i.e., the level of agreement between judges' scores).

Although the CAT demonstrates high reliability in classroom settings, there are some notable limitations in the areas of time and resources. It can be time-consuming to research, identify, and contact experts in the field that may be a good fit for an expert panel, especially when trying to differentiate these assessments for multiple students with interests in different domains. Finding and scheduling a time that accommodates schedules for all involved in the CAT can pose additional challenges. Some ways to address these challenges may include the following:

- Identify and contact experts within the community (e.g., local engineers, medical doctors, university professors, etc.) who may be willing to volunteer their time and expertise.
- Allow students to co-create their learning experiences. Depending on the grade level, research and identify experts within their domain of interest. Under this

strategy, students familiarize themselves with leading experts in the field and can practice professional communication skills.

- Engage students in creative collaborations. Group students according to their domain of interest. When students collaborate on their work, resources such as time and number of judges can be minimized.
- Use technology to your advantage. Invite experts to participate through online meeting platforms such as Wonder (www.wonder.me/), Zoom (https://zoom.us/), Miro (https://miro.com/index/).
- Tap into peer-mentoring. Collaborate with teachers in the school district and inquire about potential student experts. For instance, primary grade level teachers can recruit high school students to serve as judges on an expert panel. For secondary grade level teachers, teachers can recruit higher education students from their local community to serve on an expert panel.
- Engage students in authentic roles. Team up with a teacher in the same grade level and have students from each class judge one another's creative products, using CAT procedures.

Domain-Specific Creativity Rubrics

In addition to the CAT, teachers can assess creative products through domain-specific creativity rubrics. Rubrics are tools that explicitly define performance expectations and are used to score an assignment or artifact. Rubrics divide a product into components and explicitly describe characteristics associated with each component while also accounting for multiple levels of mastery. Overall, rubrics are a beneficial assessment method as they can be used to explicitly outline expectations and provide targeted feedback related to various aspects of a product.

When designed according to the domain, rubrics can convey expectations specific to the discipline and can be used to document advanced growth relative to specific skills for a particular domain. Creativity rubrics should consider a definition of creativity and the qualities unique to the domain. Essentially, domain-specific rubrics should acknowledge the constraints and parameters unique to the discipline and measure creative products in a way that imitates how creativity is assessed in an authentic context (Kettler & Bower, 2017). When developing rubrics to assess creative products, it is important to include criteria that target creativity skills that are unique to the domain. Creativity rubric development should follow these three general rules: (a) align the creativity rubric with the definition of creativity, (b) make a list of key components relative to the domain to be included, and (c) develop a scale and set a range of scores that focus on *growth and mastery* for each category.

The Taxonomy of Creative Thinking (see Table 7.2) may be used to guide rubric development. The Taxonomy of Creative Thinking includes five areas of creative thinking and 16 student expectations. These expectations are written in such a way that provides wide application across all subject areas and grade levels.

When using rubrics to assess creative products, it is important to develop a process that will increase inter-rater reliability. For example, Clary and colleagues (2011) found discrepancies between rubric scores and judges' top three picks of creative projects. When examining these discrepancies, they found that most of the judges did not understand how to use the rubric and, that those judges who worked with a class were more likely to give higher scores to that class. To mitigate similar problems and effectively assess creative products using rubrics, the following procedures are recommended:

- If possible, use multiple assessors who are trained in using the designed rubric.
- In addition to training, provide clear instructions to individuals who will score or rate creative products.
- As a group, discuss how scores will be assigned according to the technical language in the rubric (Clary et al., 2011).

Students can also co-create rubrics with their teachers. Co-construction of various learning experiences allows students to adopt an active role in their learning. Through ethnographic case studies of creative pedagogy, Craft et al. (2014) identified co-construction, students' control/agency and ownership in learning, and high expectations for creative engagement as essential characteristics of creative pedagogy. In the Craft et al. study, teachers provided opportunities for co-construction by allowing students to engage in the planning process for teaching and learning. Students were given the opportunity to provide feedback that related to what and how they wanted to learn about a topic. Co-constructing learning experiences also increased teachers' expectations that their students could be creative. Allowing students to co-create creativity rubrics encourages the student to research and identify creative characteristics that define their domain of interest. It also encourages students to reflect on the creative process and skills necessary to produce creatively. Moreover, students begin to learn what it means to think and act like a creative expert in the field.

Conclusion

Gifted education is the broad recognition of academic potential followed by systematic development of domain-specific talent. Creative thinking and problem solving have been staples of gifted education for decades and enhancing these skills in students with high ability prepares them for the complex work of a constantly evolving knowledge-economy (VanTassel-Baska, 2016). Domain-specific creativity assessment is complex work, requiring planning, organization, training, and administrative support. There are few shortcuts, but history tells us that what is measured and monitored will be emphasized in teaching and learning. Developing creative thinking and problem-solving skills is fundamental to talent development and exceptional academic achievement. Systematically assessing these skills for the purposes of formative assessment, student monitoring, guidance, and program evaluation should be a part of high quality, achievement-focused gifted education.

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 97–1013. https://doi.org/10.1037/0022-3514.43.5.997

Amabile, T. M. (1996). *Creativity in context: Update on the social psychology of creativity*. Westview Press.

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research on Organizational Behavior*, *36*, 157–183. https://doi.org/10.1016/j.riob.2016.10.001

Andiliou, A., & Murphy, K. P. (2010). Examining variations among researchers' and teachers' conceptualizations of creativity: A review and synthesis of contemporary research. *Educational Research Review*, *5*(3), 201–219. https://doi.org/10.1016/j.edurev.2010.07.003

Archer, E. (2017). The assessment purpose triangle: Balancing the purposes of educational assessment. *Frontiers in Education*, *2*, Article 41. https://doi.org/10.3389/feduc.2017.00041

Avitia, M. J., & Plucker, J. A. (2014). Teaching the foundations of the field: A content analysis of potential creativity texts. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(3), 378–382. https://doi.org/10.1037/a0036760

Baer, J. (1998). The case for domain specificity. *Creativity Research Journal*, *11*(2), 173–177. https://doi.org/10.1207/s15326934crj1102_7

Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review*, *37*(3), 165–178. https://doi.org/10.1080/02783193.2015.1047480

Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal*, *16*(1), 113–117. https://doi.org/10.1207/s15326934crj1601_11

Beghetto, R. A. (2006). Creative self-efficacy: Correlates in middle and secondary students. *Creativity Research Journal*, *18*(4), 447–457. https://doi.org/10.1207/s15326934crj1804_4

Beghetto, R. A., & Baxter, J. A. (2012). Exploring student beliefs and understanding in elementary science and mathematics. *Journal of Research in Science Teaching*, *49*(7), 942–960. https://doi.org/10.1002/tea.21018

Beghetto, R. A., & Kaufman, J. C. (2010). Broadening of conceptions of creativity in the classroom. In R. A. Beghetto & J. C. Kaufman (Eds.), *Nurturing creativity in the classroom* (pp. 191–205). Cambridge University Press.

Bereczki, E. O., & Kárpáti, A. (2018). Teachers' beliefs about creativity and its nurture: A systematic review of recent research literature. *Educational Research Review*, *23*(1), 25–56. https://doi.org/10.1016/j.edurev.2017.10.003

Black, P., & Wiliam, D. (1988). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148. https://doi.org/10.1002/hrm

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policies, & Practice*, *8*(2), 153–169. https://doi.org/10.1080/09695940123775

Clary, R. M., Brzuszek, R. F., & Taze-Fulford, C. (2011). Measuring creativity: A case study probing rubric effectiveness for evaluation of project-based learning solutions. *Creative Education*, *2*(4), 333–340. www.scirp.org/html/7970.html

Craft, A., Cremin, T., Hay, P., & Clack, J. (2014). Creative primary schools: Developing and maintaining pedagogy for creativity. *Ethnography and Education*, *9*(1), 16–34. https://doi.org/10.1080/17457823.2013.828474

Epstein, M. H., & Sharma, J. (1998). *The behavior and emotional rating scale: A strength-based approach to assessment*. PRO-ED.

Florida, R. (2006). The flight of the creative class: The new global competition for talent. *Liberal Education*, *92*(3), 22–29. www.aacu.org/publications-research/periodicals/flight-creative-class-new-global-competition-talent

Florida, R. (2019). Maps reveal where the creative class is growing. *Bloomberg CityLab*. www.bloomberg.com/news/articles/2019-07-09/maps-reveal-where-the-creative-class-is-growing

Florida, R., & Seman, M. (2020). *Lost art: Measuring COVID-19's devastating impact on America's creative economy*. Brookings. www.brookings.edu/wp-content/uploads/2020/08/20200810_Brookingsmetro_Covid19-and-creative-economy_Final.pdf

Grigorenko, E. L., Jarvin, L., Tan, M., Sternberg, R. J. (2008). Something new in the garden: Assessing creativity in academic domains. *Psychological Science Quarterly*, *50*(2), 295–307.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454.

Haase, J., Hoff, E. V., Hanel, P. H. P., & Innes-Ker, Å. (2018). A meta-analysis of the relation between creative self-efficacy and different creativity measurements. *Creativity Research Journal*, *30*(1), 1–16. https://doi.org/10.1080/10400419.2018.1411436

Han, K.-S. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *Journal of Creative Behavior*, *37*(2), 117–142. https://doi.org/10.1002/j.2162-6057.2003.tb00829.x

Hennessey, B. A. (1994). The consensual assessment technique: An examination of the relationship between ratings of product and process creativity, *Creativity Research Journal*, 7(2), 193–208. https://doi.org/10.1080/10400419409534524

Karwowski, M. (2013). Creative mindsets: Measurement, correlates, consequences. *Psychology, Aesthetics, Creativity, and the Arts*, *8*(1), 62–70. https://doi.org/10.1037/a0034898

Karwowski, M., Gralewski, J., & Szumski, G. (2015). Teachers' effect on students' creative self-beliefs is moderated by students' gender. *Learning and Individual Differences*, *44*, 1–8. https://doi.org/10.1016/j.lindif.2015.10.001

Karwowski, M., Lebuda, I., & Beghetto, R. A. (2019). Creative self-beliefs. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (2nd ed., pp. 396–417). Cambridge University Press.

Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 187–192. https://psycnet.apa.org/doi/10.1037/aca0000217

Kettler, T., & Bower, J. (2017). Measuring creative capacity in gifted students. Comparing teacher ratings and student products. *Gifted Child Quarterly*, *61*(4), 290–299. https://doi.org/10.1177/0016986217722617

Kettler, T., Lamb, K. N., & Mullet, D. R. (2018). *Developing creativity in the classroom: Learning and innovation for 21st century schools*. Prufrock Press.

Lamb, K. N. (2020a). Cultivating creative thinking in the classroom. In J. H. Robins, J. L. Jolly, F. A. Karnes, & S. M. Bean (Eds.), *Methods and materials for teaching the gifted* (5th ed., pp. 279–297). Prufrock Press.

Lamb, K. N. (2020b). Creativity. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: A survey of current research on giftedness and talent development* (3rd ed., pp. 137–155). Prufrock Press.

Mullet, D. R., Willerson, A., Lamb, K. N., & Kettler, T. (2016). Teacher perceptions of creativity: A systematic review of the literature. *Thinking Skills and Creativity*, *21*(1), 9–30. https://doi.org/10/1016/j.tsc.2016.001

Mumford, M. D., Martin, R., Elliott, S., & McIntosh, T. (2018). Creative thinking in the real world: Processing in context. In R. J. Sternberg & J. C. Kaufman (Eds.), *The nature of human creativity* (pp. 147–165). Cambridge University Press.

Mumford, M. D., Medeiros, K. E., & Partlow, P. J. (2012). Creative thinking: Processes, strategies, and knowledge. *Journal of Creative Behavior*, *46*(1), 30–47. https://doi.org/10.1002/jocb.003

National Association for Gifted Children (NAGC). (2019). *Pre-K to grade 12 gifted programming standards*. www.nagc.org/resources-publications/resources/national-standards-gifted-and-talented-education/pre-k-grade-12

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policies, & Practice*, *13*(2), 149–170. https://doi.org/10.1080/09695940701478321

Piirto, J. (2000). The Piirto pyramid of talent development. *Gifted Child Today*, *23*(6), 22–29. https://doi.org/10.1177%2F107621750002300608

Piirto, J. (2014). *Organic creativity in the classroom: Teaching to intuition in academics and the arts*. Prufrock Press.

Piirto, J. (2022) Talent development in the arts: A unique pathway or branches on a common road? In J. VanTassel-Baska (Ed.), *Talent development in gifted education* (pp. 195–211). Routledge.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions. *Educational Psychologist*, *39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1

Proctor, R. M. J., & Burnett, P. C. (2004). Measuring cognitive and dispositional characteristics of creativity in elementary students. *Creativity Research Journal*, *16*(4), 421–429. https://doi.org/10.1080/10400410409534553

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., Gavin, M. K., Reis, S. M., Siegle, D., & Sytsma, R. E. (2013). *Scales for rating the behavioral characteristics of superior students.* Prufrock Press.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305–310. www.jstor.org/stable/20342603

Robinson, K. (2006, February). *Do schools kill creativity?* [Video]. TED Conferences. www.ted.com/talks/sir_ken_robinson_do_schools_kill_creativity?language=en

Rubenstein, L. D., Ridgley, L. M., Callan, G. L., Karami, S., & Ehlinger, J. (2018). How teachers perceive factors that influence creativity development: Applying a social cognitive theory perspective. *Teaching and Teacher Education*, *70*, 100–110. https://doi.org/10.1016/j.tate.2017.11.012

Ryser, G. R., & McConnell, K. (2004). *Scales for identifying gifted students*. Prufrock Press.

Said-Metwaly, S., den Noortgate, W. V., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity: Theories, Research, Applications*, *4*(2), 238–275. https://doi.org/10.1515/ctra-2017-0013

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482–496. https://doi.org/10.1016/j.tate.2009.06.007

Scott, G., Leritz, L. E., & Mumford, M. D. (2004). The effectiveness of creativity training: A quantitative review. *Creativity Research Journal*, *16*(4), 361–388. https://doi.org/10.1080/10400410409534549

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal*, *45*(6), 1137–1148. https://doi.org/10.5465/3069429

Treffinger, D. J., Schoonover, P. F., & Selby, E. C. (2013). *Educating for creativity and innovation*. Prufrock Press.

VanDenBerg, J. E., & Grealish, E. M. (1996). Individualized services and supports through the wraparound process: Philosophy and procedures. *Journal of Child and Family Studies*, *5*(1), 7–22. https://doi.org/10.1007/BF02234675

VanTassel-Baska, J. (2016). Higher level thinking in gifted education. In J. C. Kaufman & J. Baer (Eds.), *Creativity and reason in cognitive development* (pp. 92–113). Cambridge University Press.

CHAPTER

8 Curriculum-Based Assessments

Tracey N. Sulak, Maryann R. Hebda,
Celeste D. C. Sodergren, and Jennifer H. Robins

Definition

Curriculum-based assessments (CBAs) offer teachers and administrators a powerful tool for connecting curriculum to student learning. Assessment, in general, is a process that uses information to make decisions about teaching and learning; this may involve individual student programming, instructional design and delivery for a class, or even campus- or district-level programming. CBAs connect the curriculum and assessment process by providing criterion-referenced information that may be used as formative or summative assessments (Macy & Bricker, 2006). In other words, CBAs offer information for designing more appropriate, targeted learning experiences for all students or for measuring the amount of learning resulting from those experiences. For gifted students, CBA assessment data can be used to create individualized or group learning experiences that address specific needs or interests. They also provide useful instructional data on student short term learning gains.

To understand the role of CBAs in the assessment process, we need to examine how they connect curriculum and assessment. A curriculum-based assessment is criterion-referenced, which means everything on the assessment can be matched to a standard (Burns, 2004). Standards serve as the backbone of the curriculum; some states have their own standards, such as the Texas Essential Knowledge and Skills (TEKS), and other states use a common set of standards, like the Common Core State Standards. When an assessment is criterion-referenced based on a set of standards, every question or concept can be connected to one or more standards. Student learning then is measured by the items, or standards, answered correctly. In this sense, CBAs may also be used at a programmatic level to evaluate program performance against standards, like those included in the National Association for Gifted Children (NAGC, 2019) Pre-K–Grade 12 Gifted Education Programming Standards. Similar to constructing a CBA for individual use, programmatic evaluations can link each NAGC standard to evaluation items that provide information for data-based decision making at the program level.

The current state of accountability in education relies on information from standardized, formal assessments to evaluate students' individual learning, teacher performance, and school performance. It is important to ensure that assessments are both valid and reliable. An assessment's validity is the extent to which an instrument measures

DOI: 10.4324/9781003285991-13

what it purports to measure. Reliability, on the other hand, focuses on how consistently the assessment

> (a)measures the same trait or construct (i.e., internal consistency), (b) yields the same score after repeated administrations (i.e., test-retest reliability), (c) generates similar scores on different versions of the assessment (i.e., alternate, parallel, or equivalent forms reliability), and (d) is scored similarly between two or more observers or raters (i.e., interrater/interscorer reliability). (Robins & Jolly, 2018, p. 78).

For an assessment to be formal, it must have standardized instructions and procedures that maintain the consistency in how the assessment is given.

Accountability measures, such as MAP and state standardized assessments, link each item to the curriculum used in the school and offer schools a method for assessing the strengths of teaching in specific areas. Advanced Placement (AP) exams function similarly. NAGC (2019) Standard 2.1.3 suggests standardized assessments may be adapted to above-level testing for students performing above grade level. In this case, the results would be used not for accountability but to serve as a way to measure growth in high-achieving students.

In addition to formal CBAs, teachers use informal CBAs, including teacher-made tests, to measure student learning for a specific unit of material. Although these assessments are considered informal because they do not have standardized instructions and procedures, creating an informal CBA requires teachers to work through specific steps, such as identifying the skills for the assessment, factoring in the level of demand found in the curriculum, and considering individual student needs and characteristics (Jones et al., 1998).

Purpose

Because CBAs are directly linked to the curriculum, the results can be used for a variety of assessment purposes that culminate in one overarching goal: identifying the content for instruction. Burns (2004) suggested that CBAs can answer questions like "what can the student do, how does the student think, how does the student approach what he or she is unsure of, [and] as a teacher, now what do I do?" (p. 65). Answering these questions help teachers understand what a student knows and does not know within the school's curriculum, select material that is challenging but not frustrating, and differentiate based on individual student needs (Macy & Bricker, 2006). Using CBAs as formative assessments allows teachers to pretest students before instruction and provide ongoing information about student performance. When CBAs function as summative, or postassessments, they provide information about students' growth or learning after instruction.

When teachers use CBAs to plan instruction, they are then able to design learning experiences for students that are within their Zone of Proximal Development (ZPD; Vygotsky, 1978), with scaffolding when necessary. The data collected tie directly to the local curriculum, giving teachers information on what a student knows and does not know at a specific point in time. By comparing the data from CBAs to a district's local scope and sequence, it becomes apparent whether a student is performing as expected, above level, or below level. This use is particularly important when considering students who are twice-exceptional (2e), because these students often excel in pockets of the curriculum but need support in other areas. Instruction that is planned through such data allow teachers to scaffold learning in areas that need support while also challenging the student in areas of strength. Similarly, information on a student's performance within the curriculum may be used to create a personalized talent development plan.

Because CBAs relate directly to the curriculum, they can be used to plan instruction for students who are not performing on grade level. Students performing at grade level should show mastery of past content and concepts but may miss items related to concepts that have not yet been taught. If students are performing above grade level, they are likely to demonstrate mastery of some skills and concepts that have not been introduced while also showing mastery of skills and concepts that have already been taught. When used in this way, CBAs help teachers teach to students' needs; skills and knowledge from outside the scope and sequence may be incorporated into instruction if it matches student needs.

Decisions about how to teach the content can also come from such assessment results. Assessment questions that require different levels of processing offer teachers insight about a student's stage of learning. For example, if a student answers basic factual questions that rely on remembering information correctly, but miss items that require deeper processing, then the student may be in the acquisition phase of learning for that concept. Students who show mastery of challenging questions that require integration of concepts may be in the generalization or adaptation phase of learning (Anderson & Krathwohl, 2001). The level of complexity of a task can be adjusted to match a student's needs. For students at an acquisition level, instructional objectives may need to be less complex and focus on learning specific skills and knowledge; in other words, students may need to practice the skills in context to develop fluency, but the complexity of the context should not interfere with developing accuracy. Once the student is fluent, the skill can be generalized to other contexts, and learning experiences can be more complex, combining multiple concepts and requiring deep processing.

In addition to decisions about content and instructional processes, CBA results can be used to support effective differentiation in pacing. Teachers need to understand the relationship between learner characteristics, learning experiences, and teaching, using feedback from these processes to maximize student performance (Jones et al., 1998). Some students will master concepts with little to no direct instruction and with few repetitions, whereas others will need more structured guidance. CBA data sometimes can illustrate the pace at which students learn and allow teachers to make decisions about curriculum compacting or within-class grouping. Adjusting groups and pacing, based on data, gives all students a chance at individualized instruction that is adjusted to their learning rate.

Using CBAs in Gifted Education

At the program and classroom levels, assessments for gifted and talented students should be appropriately challenging, intentionally designed to inform instruction, and include multiple measures of qualitative and quantitative data. The recommendations offered below provide a range of options and examples to guide implementation of best practices in gifted curriculum-based educational assessments.

Multi-Measure Assessments

Multi-measure assessments present a variety of snapshots that, in combination, form a complete appraisal of a student's growth and learning (VanTassel-Baska, 2014). This collection of assessment results is meant to incorporate informal teacher observations, student self-reflections, work samples from the classroom, and progress on standardized assessments. This provides the greatest flexibility to measure and analyze the talents of a diverse range of students. Based on the goals of the gifted and talented (GT) program, a grouping of quality curriculum-based assessments should be reviewed and selected or designed for the purpose of capturing a student profile of strengths, interests, and diversity (NAGC, 2019). Program goals should inform both the scope and level of the content and skills presented in assessments. In the process of review, the technical quality should be explored by evaluating the psychometrics and potential for bias of each assessment, in addition to aligning with the intended purpose. For an in-depth discussion of technical qualities of assessments and methods for selecting appropriate assessments, see Appendix B of the *NAGC Pre-K–Grade 12 Gifted Education Programming Standards: A Guide to Planning and Implementing Quality Services for Gifted Students* (2nd ed., Sulak & Johnsen, 2022).

To provide a complete picture of student learning, a thorough portfolio should include both formal and informal varieties of formative and summative assessments. Each of these contributes valuable input for teachers to use to make decisions about what to teach and how to differentiate. Formal, summative assessments receive significant attention in gifted education due to their use for high stakes decisions such as identification. However, even if formal assessments are valid for GT students, they only contribute one data point toward an overall understanding of student ability (VanTassel-Baska, 2014). Formative assessments are frequently used in classrooms of all kinds, whether they are informal formative checks for understanding (e.g., thumbs up/down, turn and talk, whole class samples) or more formal (e.g., exit tickets, reflective journals, assignments that are completed individually and submitted for feedback). Whether qualitative or quantitative in nature, one challenge for educators is to judge which and how many assessments should be documented to show growth. There is no rule or single correct answer; in fact, it depends on the learning context, the student, and timing.

The process of multi-measure assessments and data collection allows all stakeholders to see growth in behavior, learning, and achievement over time, as well as document that growth. Teachers, in particular, can use and interpret curriculum-based assessments to differentiate learning experiences and assist students in short- and long-term goal setting. As these goals and objectives are monitored over time using assessment data, students can track their progress and teachers can provide and adjust appropriate interventions. Patterns and trends can be established over time, which makes assessments and data meaningful for both students and stakeholders (Tomlinson et al., 2015). Technology also can be particularly useful in facilitating the collection, management, and synthesis of such data in electronic portfolios.

Intentionally Designed Assessments

Intentionally designed, standards-based assessments inform instruction and can be used as pretests or formative and ongoing assessments. Initially, assessments should be developed alongside curriculum, with combined input from professionals and educators with expertise in all applicable domains (e.g., content area, special education, gifted education, assessment; Callahan et al., 2015). Curricular standards drive both instruction and assessment, so accurate content representation and appropriate level of depth, complexity, and thought process are vital to a valid match between curriculum and summative assessment. In addition, corresponding pretests and posttests should be developed for every unit of study, exhibiting similar scope and depth of content. As gifted students complete the pretest, the resulting data should be interpreted by the teacher as to how to accelerate and enrich the upcoming unit. Pretesting should also be a schoolwide practice, as it can be a shock for some students to be "tested" over content they have not yet been taught. Therefore, teachers must be prepared for a variety of responses to pretests and must be intentional about communicating to students and parents their formative and ungraded nature.

Ongoing formative assessments take many forms in classrooms (e.g., verbal, written, virtual) and inform instruction for each lesson in each disciplinary or multidisciplinary subject. These quick checks for understanding mainly inform the daily adjustments to instruction that teachers make in response to student needs, such as pacing, depth, and complexity. The purpose of this type of CBA for gifted students is not to receive a grade, but rather to receive useful feedback and facilitate reflection and planning for growth (Tomlinson et al., 2015). For educators, these quick checks allow patterns to emerge among groups and individuals, which should be interpreted to shape future instructional strategies and content focus. Especially for advanced learners, assessment itself should be seen as a learning opportunity (Varsavsky et al., 2013). When educators use formative assessments to provide quality feedback, students can be more aware about their own strengths and gaps, revealing meaningful learning opportunities throughout the cycle of assessment (Tomlinson et al., 2015), which assists in the development of metacognition.

Appropriately Challenging Assessments

Appropriately challenging assessments should match the differentiated content and strategies of curriculum and instruction. Just as the curriculum requires GT students to practice higher order thinking, so should the assessment (VanTassel-Baska et al., 2002). To validly measure the intended skills and content, an open-ended assessment requiring students to interpret, transpose, infer, or reshape concepts is more appropriate for advanced learners (Purcell et al., 2002). This aligns with a tiered model of differentiated learning and assessment, designed to measure individual growth (Brown, 2012). As educators and specialists are interpreting assessment results, this feedback and growth component is particularly nuanced for GT learners. For example, if one student consistently earns full marks on criterion-referenced tests, a ceiling effect—or insufficient challenge—could exist in the curriculum and the assessment and resulting scores cannot reflect growth over time. When a situation like this occurs, other curriculum opportunities must be explored for the student.

One option that has potential to better show growth in learning is using standardized off-level assessments for GT students (Brown, 2012). This approach can present challenges in terms of mandatory grade-level standardized assessments; however, if it is appropriate for the student, both on- and off-level assessments could be pursued.

Examples of CBAs Used in the Classroom

As gifted programs implement more project-based curricula, the need for quality performance-based assessments increases accordingly. This type of curriculum-based assessment calls for an element of student choice, but that customization results in unique insights into the student's true capability in the domain. VanTassel-Baska (2014) suggested creating open-ended, advanced-level tasks that require each student to clearly articulate their thinking process.

Social Studies and Reading Example

When assessing the Common Core State Standards (CCSS) in Literacy within a middle school U.S. History class, a performance-based approach would be a strong match for assessing "aspects of a text that reveal an author's point of view or purpose (e.g., loaded language, inclusion or avoidance of particular facts)" (CCSS.ELA-LITERACY.RH.6–8.6). Above-level texts such as primary documents from the Civil War era could be analyzed by gifted students using biographies, archival records including personal correspondence, and local newspaper clippings. Any number of creative products could address the goals of this standard (i.e., narrated slideshow, speech from the author's perspective, written analysis) and provide the opportunity for gifted students to present the full depth of their understanding. The use of a quality research rubric for advanced products (see Figure 8.1) keeps the focus on the depth and complexity of the student's thinking about curricular standards while minimizing ceiling effects (Hughes et al., 2014).

Science Example

Science also presents a natural fit for a performance-based approach to assessment. Problem solving is at the heart of scientific inquiry, so open-ended assessments align exceptionally well with the Next Generation Science Standards (NGSS). Designing experiments, software, databases, models, and simulations are all examples of authentic assessment approaches for science and STEM topics (Sedivy-Benton et al., 2016). For example, the Fowler Science Process Skills Assessment presents the opportunity for students to design and write an experiment as a preassessment (Are bees attracted to diet soda?) and to replicate the process as a post assessment (Are worms attracted to light?) with the presence of experimental design concepts indicating growth over time (Fowler, 1990). This rubric for the assessment is purposefully designed to inform teachers of each student's growth from pre- to post assessment, so it could easily be supplemented with shorter, student-friendly rubrics as ongoing formative assessment between the pre- and post-assessment. (For more information, please visit https://education.wm.edu/centers/cfge/_documents/curriculum/science/dietcolatest.pdf.)

The most effective rubrics provide useful feedback for improvement and some description for the score (Sedivy-Benton et al., 2016; Tomlinson et al., 2015). Along with sufficiently challenging levels for gifted students to seek, performance-based rubrics are a valuable tool for students to see their current achievement and set new performance goals. The Fowler rubric is very helpful in this regard as it can be used across years as a time series approach to growth in the scientific research process.

Math Example

Compacted or accelerated curriculum is appropriate for many gifted students in math. For placement level decisions to be made, some schools and districts choose from existing traditional aptitude assessments based in CCSSM. Some assessments are useful beyond the initial placement and can be implemented for continued monitoring of above-level progress, such as the Northwest Evaluation Association's Measure of

Element	0: Poor	1: Acceptable	2: Good	3: Exemplary
Use of advanced content	No advanced content integrated into product.	Some advanced material present with integration into the product.	Advanced material present and well integrated into product.	Significant quantity and quality of advanced materials integrated into products.
Critical thinking	Insights are shallow or reflect only surface-level understanding.	Insights are reflective and make connections with content.	Insights are in-depth and connect content in thoughtful and thorough manners.	Insights add considerably to the understanding of a piece and provide new learnings and new understandings.
Depth	Products reflect only single-level or surface-level understanding. There are no or insufficient connections to other ideas and/or content.	Products reflect multiple levels of understanding. There are some relevant connections to other ideas and content.	Products reflect understanding of the concept and how it relates to other concepts. Sufficient and insightful connections to other ideas are made.	Significantly insightful connections to other ideas, concepts, and/or content is made. Ideas are explored and boundaries of understanding defined.
Creative thinking	Products or ideas are similar to those of other students and do not demonstrate an elaborated response.	Products or ideas reflect some individuality and elaboration beyond the ordinary.	Products or ideas are innovative and reflect significant individuality.	Products or ideas are advanced, unique, and innovative.

FIGURE 8.1 Assessment rubric for gifted and advanced learners

Academic Progress (MAP). This example of adaptive assessment uses item response theory to adjust difficulty based on the student's responses (https://nwea.org), which yields results that assist educators in identifying not only what has been mastered (i.e., grade-level standards), but also what each student is ready to learn next. Additional above-level math assessments that are appropriate for gifted students include traditional measures of achievement like the Woodcock-Johnson and Wechsler Intelligence Scales, which are more widely available due to use in special education and other educational applications (for more examples, see Johnsen et al., 2014).

The NAGC 2019 Pre-K–Grade 12 Gifted Education Programming Standards clearly state that educators should use a variety of qualitative and quantitative assessment data to guide instruction and interventions. As a school subject, mathematics traditionally offers quantitative skills, but for gifted students it should still include the creativity, depth and complexity, and multidisciplinary applications that are best practices for all content areas (Johnsen et al., 2014). For example, if a pretest revealed that several gifted elementary students had already mastered fractions, general education and gifted teachers could collaborate to accelerate the group to ratios. As a project, students may interview an architect, draft a blueprint, and 3D print a model at several scaled sizes. Summative assessment might include a group presentation of how their project reflects real world applications, math journal entries explaining their calculations, and a posted video response revealing their thought processes. By incorporating qualitative components in mathematics assessment, educators can better judge not only mastery, but also a student's interactions with skills and concepts.

Future Research and Directions for Practice

As noted above, curriculum-based assessments help connect a district's curriculum to student learning. This topic is ripe for future research and directions for practice. Plucker and Callahan (2014) noted that it would be beneficial to have more empirical studies of CBAs based on gifted education curricular models. More empirical research in this area would support administrators and educators in using alternative assessment approaches to determine gifted student learning.

In addition, there is evidence that suggests that prescriptive curricula that is specifically developed for advanced learners—based on information from CBAs that indicate these students are above level—often allow for more advanced student outcomes (Plucker & Callahan, 2014). It is important to note that

> These pre-differentiated curricula are characterized by clearly stated learning outcomes, strategies for formative assessment accompanied by direction on how to use data effectively in modifying curricular and instructional decisions, lesson plans that reflect the modification based on student data, and assessments that reflect the varied learning levels. Both standardized achievement tests and specific curricular assessments have shown that students who were randomly chosen to participate in lessons using these curricula achieved at higher levels than did similar students who were offered standard lessons. (Plucker & Callahan, 2020/2021, p. 18).

Educators can use the data obtained from CBAs to modify both curriculum and instruction for their students.

In addition, administrators should provide professional learning opportunities for teachers about CBAs, RtI, and gifted students. Educators must have a deep understanding of all three of these areas in order to ensure that they are meeting the needs of all students, including those who are gifted. Educators who use data obtained from CBAs need to change their instruction to make sure each student is learning something new every day, no matter their level.

Conclusion

Gifted and talented students possess a variety of unique strengths and needs, and like these students, their assessments must comprise variety. As professional educators, specialists, and administrators monitor and select appropriate assessments for GT students, CBAs offer unique insights into student learning and rates of growth. As more educators and school systems adopt curriculum-based assessments as a universal screener for gifted education, it would be expected that a more diverse group of learners would be identified for more challenging instruction (Dixson et al., 2021). CBAs offer greater access to advanced learning opportunities and provide a more equitable alternative to identifying and serving children with advanced abilities. By tying assessment to instruction, educators can ensure continuous and ongoing instruction that is aligned to each learner's ability.

References

Anderson, L., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abridged ed.). Allyn and Bacon.

Brown, E. F. (2012). Is response to intervention and gifted assessment compatible? *Journal of Psychoeducational Assessment*, *30*(1), 103–116. https://doi.org/10.1177/0734282911428200

Burns, M. K. (2004). Using curriculum based assessment in consultation: A review of three levels of research. *Journal of Educational and Psychological Consultation*, *15*(1), 63–78. https://doi.org/10.1207/s1532768xjepc1501_3

Callahan, C. M., Moon, T. R., Oh, S., Azano, A. P., & Hailey, E. P. (2015). What works in gifted education: Documenting the effects of an integrated curricular/instructional model for gifted students. *American Educational Research Journal*, *52*(1), 137–167. https://doi.org/10.3102/0002831214549448

Dixson, D. D., Peters, S. J., Makel, M. C., Jolly, J. L., Matthews, M. S., Miller, E. M., Rambo-Hernandez, K. E., Rinn, A. N., Robins, J. H., & Wilson, H. E. (2020–2021). A call to reframe gifted education as maximizing learning. *Phi Delta Kappan*, *102*(4), 22–25. https://doi.org/10.1177/0031721720978057

Fowler, M. (1990). The diet cola test. *Science Scope*, *13*(4), 32–34.

Hughes, C. E., Kettler, T., Shaunessy-Dedrick, E., & VanTassel-Baska, J. (2014). *A teacher's guide to using the Common Core State Standards with gifted and advanced learners in the English Language Arts*. Prufrock Press.

Johnsen, S., Ryser, G., & Assouline, S. (2014). *A teacher's guide to using the Common Core State Standards with mathematically gifted and advanced learners*. Prufrock Press.

Jones, E. D., Southern, W. T., & Brigham, F. J. (1998). Curriculum-based assessment: Testing what is taught and teaching what is tested. *Intervention in School and Clinic*, *33*(4), 239–249. https://doi.org/10.1177/105345129803300407

Macy, M. G., & Bricker, D. D. (2006). Practical applications for using curriculum-based assessment to create embedded learning opportunities for young children. *Young Exceptional Children*, *9*(4), 12–21. https://doi.org/10.1177/109625060600900402

National Association for Gifted Children. (2019). *2019 Pre-K–Grade 12 Gifted Education Programming Standards*. www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20 Standards.pdf

Plucker, J. A., & Callahan, C. M. (2014). Research on giftedness and gifted education: Status of the field and considerations for the future. *Exceptional Children*, *80*(4), 390–406. https://doi.org/ 10.1177/0014402914527244

Plucker, J. A., & Callahan, C. M. (2020–2021). The evidence base for advanced learning programs. *Phi Delta Kappan*, *102*(4), 15–21. https://doi.org/10.1177/0031721720978056

Purcell, J. H., Burns, D. E., Tomlinson, C. A., Imbeau, M. B., & Martin, J. L. (2002). Bridging the gap: A tool and technique to analyze and evaluate gifted education curricular units. *Gifted Child Quarterly*, *46*(4), 306–321. https://doi.org/10.1177/001698620204600407

Robins, J. H., & Jolly, J. L. (2018). Technical information regarding assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 73–116). Prufrock Press.

Sedivy-Benton, A. L., Olvey, H. A., & Van Haneghan, J. P. (2016). Assessing aptitude and achievement in STEM teaching and learning. In B. MacFarlane (Ed.), *STEM education for high-ability learners: Designing and implementing programming* (pp. 123–138). Prufrock Press.

Sulak, T. N., & Johnsen, S. K. (2022). Appendix B. Assessments for measuring student outcomes. In S. K. Johnsen (Ed.), *NAGC Pre-K–Grade 12 Gifted Education Programming Standards: A guide to planning and implementing quality services for gifted students* (pp. 344–378). Routledge.

Tomlinson, C. A., Moon, T., & Imbeau, M. B. (2015). *Assessment and student success in a differentiated classroom* [White Paper]. www.ascd.org

VanTassel-Baska, J. (2014). Performance-based assessment: The road to authentic learning for the gifted. *Gifted Child Today*, *37*(1), 41–47. https://doi.org/10.1177/1076217513509618

VanTassel-Baska, J., Zuo, L., Avery, L. D., & Little, C. A. (2002). A curriculum study of gifted-student learning in the language arts. *Gifted Child Quarterly*, *46*(1), 30–44. https://doi.org/10.1177/ 001698620204600104

Varsavsky, C., & Rayner, G. (2013). Strategies that challenge: Exploring the use of differentiated assessment to challenge high-achieving students in large enrolment undergraduate cohorts. *Assessment & Evaluation in Higher Education*, *38*(7), 789–802.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press. https://doi.org/10.1080/02602938.2012.714739

CHAPTER

9 Performance-Based Assessments for Secondary Gifted Students

Elissa F. Brown

Designing assessments that take into account cognitive and personal dimensions is critical in recognizing in advanced and gifted secondary learners strengths while providing tasks that further their thinking and problem-solving capabilities and utilize positive personal traits. Designing such assessments is labor intensive, but it yields benefits for enhancing instruction and student growth (VanTassel-Baska, 2008). This chapter provides descriptions and discussion about these types of performance-based assessments: (a) traditional curriculum-based assessments, (b) above grade level content-based assessment such as AP and IB, (c) problem-based assessments, (d) portfolio-based assessments, and (e) the role of competitions. This approach to assessment at secondary levels is already in place and used in various forms from teacher-developed tools to advanced programs' standardized templates. Examples of these types of assessments follow along with caveats related to their use.

Importance of Using a Mix of Assessment Types with Gifted Adolescents

Abilities may become more differentiated in gifted adolescents by area (i.e., verbal and quantitative) as well as in the level of talent (i.e., how advanced a skill level or ability is) (Dixon & Moon, 2015; Subotnik et al., 2011). Therefore, educators may assume that high school gifted adolescents are capable of thinking, learning, and demonstrating their learning using assessment measures that include a range of levels from a college-age student to a professional in a field (Dixon & Moon, 2015). Providing one type of assessment, such as a summative standardized state test, only provides one data point on a prescribed set of content items. Incorporating a more robust approach to assessment with a mix of traditional and performance-based assessments allows gifted adolescents to show variations in learning gains and provides evidence of the level of performance in a given domain contextually. The Common Core standards and their commensurate assessments' intent are designed to ensure that students graduating from high school are prepared to

DOI: 10.4324/9781003285991-14

take credit-bearing introductory courses in two- or four-year college programs or to enter the workforce. The standards are benchmarked against the Advanced Placement program as well as national and international frameworks: the Third International Mathematics and Science Study (TIMSS), Program for International Student Assessment (PISA), and the National Assessment of Educational Progress (NAEP; Wiley et al., 2010). Gifted and talented students typically grasp curriculum concepts more quickly and deeply than their age peers. They achieve grade-level expectations earlier than specified in state and national standards and generally need far fewer instructional and practice experiences to achieve mastery. To make continuous progress in their areas of talent, gifted students need learning experiences that extend and enrich the standards and require students to apply complex, creative, and innovative thinking to authentic problems. To measure their progress, they need assessments that can capture the level of learning achievement that they can produce (VanTassel-Baska, 2013).

The purposes of assessments vary. Some are externally driven, such as state content tests administered at the end of the year and used by policymakers and school districts to determine subsequent placement in courses or school report cards that are shared with the public. Some are linked to course credit, class rank, and accrue college credits or placement, such as Advanced Placement or International Baccalaureate, and still other assessments are those that teachers create. These teacher-developed formative assessments provide insights into ongoing student progress in a domain and are used for curriculum planning, assessing process skills such as writing, articulation of ideas and reasoning, or to gauge student understanding of specific assignments. Olszewski-Kubilius and Thomson (2014) noted that the use of multiple assessment measures (traditional and performance-based) enhance understanding of how to counsel students and parents in the talent development process.

Traditional Curriculum-Based Assessments

More traditional forms of assessment such as state tests may not include items that allow students to exhibit advanced thinking or growth over time in a domain because of ceiling effects (VanTassel-Baska & Baska, 2019). These assessments are administered annually to measure student achievement of state standards at specific grade levels. These assessments can show students' acquisition of knowledge and skills as compared to peers in the same age group or grade level and whether or not they have met state standards, which are required for all public school students.

In gifted programs, these types of assessments are frequently used to examine advanced learning even though they are not useful in making judgments about individual gifted student's learning progress because of ceiling effects. Abrams et al. (2003) found that state-mandated testing programs can even lead to instruction that contradicts teachers' views of sound educational practice. Teachers reported that the pressure to raise test scores encouraged them to emphasize instructional and assessment practices that mirrored the content and format of the state test.

In summary, traditional state assessments are used to determine student's acquisition of knowledge and skills that are related to grade-level standards. These assessments may be helpful in examining students' progress within the general education program; however, they do not have sufficient ceiling to measure the progress of gifted students who are advanced beyond their current grade level.

Performance-Based Assessments

Performance-based assessments require students to construct a response, solution, or product through focusing on assessment *for* learning (i.e., formative assessment) as opposed to assessment *of* learning (i.e., summative assessment; Chappuis & Stiggins, 2017). They provide an alternative way of looking at student learning and can complement a comprehensive assessment approach for gifted secondary students. Broadly speaking, performance-based assessments improve student learning because they assess the effectiveness of the interaction between students' current knowledge and ability and teacher instruction. Two of the biggest differences between performance-based and traditional measures are context and purpose. In traditional measures, the context is not considered because the assessment typically is separated from the learning activity. Standardization is important to ensure reliability and validity. In contrast, performance-based assessments are embedded in learning experiences, which occur in the classroom. Their reliability and validity are derived from the alignment of the activity to curricular objectives and from the teacher who aligns the activity to the students' strengths and needs for improvement (Abeywickrama, 2012). In addition, one of the main purposes of traditional summative measures is to assess learning in a specific body of content, and the focus is on the outcome of a program (assessment *of* learning). Whereas a major purpose of performance-based assessment is that it can assess the capacity for student growth. VanTassel-Baska et al. (2009) have reported that the use of performance-based assessments across content areas with gifted students provides learning gains in scientific research skills, literary analysis and interpretation, and persuasive writing.

Shepard (2001) cited the following characteristics of performance-based assessments:

- The teacher is involved from beginning to end in assessment activities.
- A number of samples of student work are collected over time.
- The teacher can modify and adapt the assessment to match the teaching and learning goals of the class and the diversity of the students being assessed.
- Assessments are carried out in classrooms.
- Assessments involve students more actively in the assessment process.
- Assessments allow for immediate and ongoing feedback to the students.
- Assessments stimulate continuous evaluation of teaching and learning.
- Assessments complement others forms of assessments, including state assessments and above grade-level content-based assessments.

Wiggins (1990) made the case for employing more authentic assessments over thirty years ago, stating that these assessments directly examine student performance on worthy intellectual tasks. He summarizes the reasons why authentic assessments should be used over traditional assessment measures, highlighting that they require students with acquired knowledge to be effective performers, to respond to an array of high level tasks, to craft thorough and polished products, and to respond to ill-structured demands as a professional in the field might do. Traditional tests, on the other hand, tend to require students to demonstrate their knowledge through single written responses that lack preparation or planning in the final product and that use predetermined problems rather than real ones.

Bransford et al. (2000) found that students could not learn complex problem solving at school unless they had substantive prior knowledge. They thought a much stronger way to look at whether learners have enough information is to apply what they learned in a new situation. Therefore, assessing gifted learners at the high school level to determine learning impacts needs to be revised in ways that nurture aptitudes, promote growth, and further problem-solving abilities. In short, performance-based assessments measure students' abilities to apply skills, knowledge, and infuse emotional attributes learned from the study of organized content. Typically, the task challenges students to use higher-order thinking skills to create a product or complete a process (Chun, 2010). Content-based tasks can range from a simple constructed response to a design proposal for a sustainable neighborhood. Arguably, the more authentic assessments mirror the responsibilities of a professional (e.g., engineer, consumer advocate, architect, etc.), the better the outcomes for student thinking and understanding. Performance-based assessments represent a strategic and indispensable way to assess gifted learning in a content domain.

Above-Grade-Level Content-Based Performance Assessments

Assessments that are above grade level or designed to measure advanced levels of knowledge as well as exceptional reasoning abilities are more likely to capture a student's abilities in comparison to grade-level tests that do not have enough difficult items (e.g., low ceiling). Employing off-level or above-grade-level assessments is an important component of a robust secondary assessment system since gifted learners are cognitively advanced in different domains. Using advanced content and off-level content assessments is also one form of acceleration (Assouline et al., 2015).

In most American high schools, Advanced Placement (AP) and International Baccalaureate (IB) are the most widely used secondary models for both service delivery and assessment of gifted students (Foust et al., 2008). However, in order to be placed in AP or IB classrooms, students typically are not formally identified as gifted but rather take pre-requisite courses, receive teacher recommendations to take available courses during scheduling, or perform highly on the PSAT to indicate readiness. Both AP and IB offer students the opportunity to earn college credit although program designs and requirements differ by subject area. They both include end-of-course above-grade-level exams, which many universities recognize for college credit or advanced standing. However, more than a half million low income and students of color, who would otherwise benefit if they participated at the same rate as their more advantaged peers, are missing from AP or IB courses (Theokas & Saaris, 2013). In addition, some researchers suggest that the strong focus on academic skills and test performance in the AP and IB programs may be too narrow to identify qualified students (Suldo, et al. 2018).

Advanced Placement

Advanced Placement (AP) is a program administered by the College Board that allows high school students to take designated courses that can earn a student college credit and/or qualify them to place in advanced classes as they begin college (College Board, 2015).

AP classes were created in the mid-1950s as a response to the widening gap between secondary school (high school) and college. A pilot program in 1952 had 11 subjects where high school seniors studied college level material; however, the program was not formalized until 1956 when the College Board took over and named it the College Board Advanced Placement Program. This program expanded rapidly over the years and offered a well-calibrated course of study with college-level introductory courses at the high school level, often taught by high school faculty who had received advanced training in their subject area.

Today approximately 3 million students in the United States take AP exams every year in 38 subjects. Additionally, gifted students commonly take multiple AP classes over the course of their high school careers. In terms of assessment, AP classes prepare students to take the AP exam. These standardized exams are designed to measure how well a student has mastered the content and skills of the course. Some of these exams are performance-based (e.g., English Literature, European History, Environmental Science, Statistics) while others are more traditional in orientation, using multiple choice formats and questions that test specific content acquisition (e.g., Calculus, Chemistry, Physics). AP courses and exams have become a key marker in high school coursework rigor and an integral part of the high-school-to-college transition.

AP Exams. Development committees, comprised of secondary teachers and college professors, designed AP examinations for each of the disciplines in consultation with statisticians and psychometricians to meet accepted standards for technical quality. Thus, the development process incorporated the judgments of both disciplinary and psychometric experts.

The team approach is used to score the exams as well, comprised of secondary teachers and university professors in the content discipline tested. Rating guides were developed, with scores ranging from a low of 1 to a high of 5. Technical adequacy was established for the scoring process, demonstrating both content validity and interrater reliability.

Each AP exam has its own format. On the literature exam, students are typically given a prompt to read, which is followed by a series of questions requiring responses. In the example in Appendix A, students are asked to respond to the following:

- Respond to the prompt with a thesis that presents a defensible interpretation.
- Select and use evidence to support your line of reasoning.
- Explain how the evidence supports your line of reasoning.
- Use appropriate grammar and punctuation in communicating your argument.

In order to respond and receive a score of 3 or higher, even gifted students will need to have practice with this type of writing. Historically, there was large consensus among institutions of higher education that a score of 3 or greater merited college credit. Implications for being awarded credit or skipping introductory courses and entering higher level courses (e.g., placement) have traditionally included financial savings, opportunities for additional college coursework, early graduation, the ability to pursue a double major or a combined degree, and time to study and travel aboard. Increasingly however, the assumption that a score of 3 would result in college course credit has been decreasing. Many institutions have raised the score to 4 or 5, and some are not granting credit or exemption from introductory courses at all (Schneider, 2009).

The overall benefits of AP, however, continue to be strong, particularly with traditionally underrepresented gifted populations. Gifted students from poor, immigrant, and working-class backgrounds can save semesters of tuition since all state university systems award college credit for AP courses. Federal law provides money to help states expand their AP programs and cover test fees for low-income students, and the Department of Education's Office for Civil Rights collects data to monitor participation and success rates by race/ethnicity of students. Dozens of states have sponsored AP distance learning programs to reach students in schools that don't offer AP courses, and many have invested state dollars to encourage and reward successful participation.

College Board (2013) found that strong AP programs in high schools coupled with strong AP policies at colleges, supported five major outcomes for students. Students who earned credit and advanced placement for a relevant introductory college course:

- Performed well in subsequent college courses in the discipline;
- Were more likely to major in their AP subject or a related discipline;
- Took more—not less—college course work in the discipline;
- Were more likely to graduate within five years; and
- Developed an interest in STEM subjects that led to a STEM major in college.

AP exams constitute one type of off-level content assessment since the goal of AP is to give students an introduction to the content of a college-level class. Both the content of those courses (e.g., A.P. Literature, A.P. Calculus) and their commensurate exams are considered above-grade-level curriculum and assessments. An example of an AP literature student prompt question from 2020 can be accessed here: https://apcentral.collegeboard.org/pdf/ap20-english-literature-and-composition-f2-frq-prompt-johnson.pdf

One of the distinct characteristics of Advanced Placement is the explicit link to an equivalent college-level course. Additionally, students who receive credit for introductory courses, based on their AP assessment score, have been compared in subsequent courses to those who took the college introductory courses only; the AP students performed higher, based on grades, than their peers (Hargrove et al., 2008; Keng & Dodd, 2008). The above level content "delivered" to secondary students is a form of acceleration which has a longitudinal research base of support in the field of gifted education (Colangelo et al., 2004; Assouline et al., 2015), and therefore is viewed as a viable option for gifted students who are defined by their high academic or intellectual capabilities.

International Baccalaureate

In the 1960s, Robert Leach organized a conference of internationally-minded schools, where the term "International Baccalaureate" was first mentioned. The conference was convened for teachers of social studies in international schools, who advocated for a globally-minded education. By the late 60s, teachers from the International School of Geneva developed a curriculum for students who moved internationally and who wanted to be able to attend a university anywhere in the world. The International Baccalaureate was recognized as an official organization in 1968, with a pilot program of 12 schools in 12 different countries. Today, there are over 1.25 million students in more than 100 countries worldwide who attend schools that are recognized as International Baccalaureate Organizations. Students can earn a certificate for passing the IB exam in a single subject or they can earn a full IB diploma if they pass six subjects; pass a course in Theory of Knowledge; write an "extended essay"; and complete "creativity, action, and service" hours outside of class.

IB Assessments. IB students' examinations are marked by external IB examiners (similar to AP), and the marks awarded for each course range from 1 (lowest) to 7 (highest). The International Baccalaureate program for high schools' (called the "Diploma Program") core focus remains on developing inquisitive, compassionate, and knowledgeable young people who are primed to become lifelong learners and globally-minded citizens.

The culminating assessments for IB are comprehensive because they assess a semester or year of content and embed higher level skills that require students to problem solve and think critically in their responses. The assessments are performance-based, called "constructive response" rather than multiple choice. IB examinations are designed in a manner similar to those that might be used by teachers who have a sophisticated understanding of their discipline. Thus, the development of IBO assessments depends heavily on the expertise and professional judgment of master teachers and less on psychometric calibrations per se. An example of an IB question in IB History is:

> *To what extent did World War II lead to women in the United States becoming permanent participants of the labor force?* In the student response, students will identify and evaluate sources (both primary and secondary sources), conduct an investigation of the context, and derive a reasoned conclusion consistent with the evidence and arguments, (https://ibo.org/programmes/diploma-programme/assessment-and-exams/).

The differences between the two programs' assessments (i.e., AP and IB) should be considered in light of the ways students' examination scores will be used and the kinds of inferences users expect to support with the scores. Research conducted by Dougherty et al. (2006) provided evidence that students who take AP and IB courses and score well on their commensurate assessments perform well or better in college than students who do not. Regardless of which program a secondary school implements, both AP and IB assessments are appropriate for gifted adolescents because they measure advanced content understanding, evidence-based argument in writing, and critical thinking.

Finland Example of a Performance-Based Assessment

Since, the assessment system of Finland is based around improving instruction, the majority of their assessments are formative or used to improve instruction and learning. Student assessment in Finland takes place in three arenas: within classroom practices, as the final comprehensive assessment of student progress at the culmination of basic education, and during the matriculation examination to serve as a criterion for college admission. Further, the national curriculum is evaluated using an external evaluator and data from a national standardized assessment. Teachers and schools use self-evaluation to improve education locally.

The following example comes from a summative secondary mathematics assessment in Finland where students are given "standard of living" data to solve a problem:

> *In a society the growth of the standard of living is inversely proportional to the standard of living already gained, i.e., the higher the standard of living is, the less willingness there is to raise it further. Form a differential-equation-based model describing the standard of living and solve it. Does the standard of living rise forever? Is the rate of change increasing or decreasing? Does the standard of living approach some constant level?* (Finnish National Board of Education, 2004).

This example provides the level of problem-solving that gifted secondary math students might be expected to demonstrate after having a course in calculus. Thus, performance-based assessments are seen as an effective measure of intellectual achievement or ability because they require students to demonstrate their deep understanding of content, higher-order thinking, and complex problem solving through the performance of exemplary tasks.

Social Studies High School Classroom Example of a Performance-Based Assessment

This high school performance-based assessment in social studies for gifted secondary students, undergirded by content, engages students in high end problem solving, creativity, and individual preferences. Ms. L. teaches in the social studies department in a specialized high school for gifted students in a large urban district in the northeast United States. She has been teaching at the school for over 10 years. She typically has 28 students in the last required social studies class, which is an 11th grade integrated Global/American history course covering post WWI to the present. Her final assessment for Unit 5 responds to the intent of the CCSS ELA-Literacy standards for articulating key ideas and details, craft and structure, integration of knowledge and ideas, and text complexity (www.corestandards.org/ELA-Literacy/RH/11-12/) as well as key components of best practices for assessing student growth at the secondary level. Her assessment is as follows:

> **Assignment**: *In the 1950s The United States experienced demographic, economic, and social changes as Americans returned from the war, and tried to return to a sense of normalcy. At the same time Americans were forced to reconcile with a more globalized world, and a Cold War that challenged on both an ideological front, as well as a cultural and economic one.*
>
> **Depict visually your perceptions of the condition of domestic America in the 1950s.** *You may draw, make a political cartoon, create a diagram, combine visuals and text, etc. Be creative. Be sure to convey your knowledge of the material and use your visual to communicate an argument about whether the 1950s was about conformity/homogeneity or about fault lines/dissension in American society. In essence, is there a tension between America's Cold War stance and what is being understood and lived in America?* ***Attach a written summary (1 page maximum)*** *as accompaniment to your visual to make sure that your viewer can understand your perceptions.*
>
> *In addition to the readings assigned to you for homework and during class, you may also want to look over the Nash readings on postwar America* (Nash et al., 2009).

The following key elements undergird Ms. L's assignment and align with the characteristics of performance-based assessment:

- Student choice undergirded by advanced content knowledge,
- Creativity of product design with carefully designed criteria,
- A well-contextualized problem that targets high level skills and allows students multiple access approaches, and
- Conceptual thinking about themes and issues promoted.

Darling-Hammond (2010) asserted that high-quality assessments should "emphasize deep knowledge of core concepts within and across the disciplines, problem solving, collaboration, analysis, synthesis, and critical thinking" (p. 3) incorporating "more analytic selected-response and open-ended items than many U.S. tests currently include" (p. 8). In the summative performance-based case study example, Ms. L. clearly incorporates such core content knowledge, higher level skills in thinking, and higher level concept development through her assessment.

Conditions and Strategies for Implementing Performance-Based Assessments

When developing performance-based assessments for classrooms, there should be a balance between carefully designed criteria (forming the basis of judgment) in the assessment for all students and allowance for student individuality. Additionally, beyond the specific assessment, other systems should be in place that encourage and promote student responses. When interviewed, Ms. L. shared that the following conditions were in place as she implemented the task assessment:

1. Be willing to work with students from beginning to end and have regular check-ins with students to conference with each of them.
2. Provide an "out" option, which is an important feature for gifted students who may struggle to think creatively. For example, a teacher can offer an optional activity such as writing a paper or a test with a key.
3. Throughout the unit of study, focus on the theme and include it in a summative performance-based assessment. For example, in this unit students choose a theme they understand such as power and can develop it through mini themes they have studied. Modeling thematic thinking throughout the unit of study (e.g., economic themes, political themes) also allows students to understand perspectives and diverse points of view more readily.
4. In class, implement smaller micro tasks that mirror aspects of the summative assignment. For example, share political cartoons, advertisements, or editorials of the time and have students re-create them based on a modern day issue.
5. To support their creativity, be open even if they approach the assessment in a different way from how you would have thought about it.
6. Provide space the day the assignment is due to let students see each other's work. Have them do a gallery walk or something similar, where they see connections to other works. This provides another platform for authenticity.

Strategies for Implementing Performance-Based Assessments

There are some core strategies that enable educators to cover content and move toward the use of performance-based assessments as demonstrated in the classroom example above.

These suggested core strategies are as follows:

Core Strategy #1: Preassessment and Diagnostic Follow-up. The use of preassessments at targeted points in the semester or year allows students to demonstrate prerequisite knowledge in a specific content area. For example, in a science class a preassessment could be used to determine if students know all the variables in setting up an experiment. The key to implementing preassessments though is the degree to which a teacher uses the resulting data to modify future content and instruction for data-driven instructional improvement. Just using a preassessment with no follow-up is an inefficient use of time and defeats the purpose of a valuable use of student information. One example might be to use the Fowler Test as a pre-assessment for scientific research design and then organize students in teams based on the results. A teacher in a gifted class might select the processes associated with "interpretation" as the basis for subgrouping for a week, then switch groups based on scores on "reporting" for another. Conferences would be held with each group to ensure learning progress on relevant tasks.

Core Strategy #2: In-Class Power Tasks. Use of performance-based tasks to stimulate interest and monitor readiness for a novel task is an important instructional approach. Having students apply their learning or transferring knowledge in a demonstrative way deepens student understanding (Earl, 2003). Immediate feedback through oral discussion and informal mechanisms such as observation heightens the opportunity for authentic learning to occur. For example, having students write a persuasive essay in response to an authentic prompt within a time frame teaches students to be succinct, cogent, and yet comprehensive in their writing.

Core Strategy #3: Specific Feedback for Improvement. For students to improve as thinkers, writers, technicians, and creative individuals, they need advice, counsel, tutoring, and feedback on how to progress and improve. Coaching may be required to encourage such progress. Peer or teacher feedback can be beneficial as well as students judging themselves against an external standard to target specific areas for improvement. To facilitate learning, educators need to provide frequent opportunities for students to test their knowledge for understanding and subsequently get feedback for revision (Karpov & Bransford, 1995). Another approach would be to teach students the rubric for assessing their learning in an area and have them apply it to their own and their classmates' work. Continuing to use the same rubric over time provides a basis for discussing progress.

Core Strategy #4: Product Development. Regardless of the product, having students demonstrate their understanding in some format, even through applied homework, enhances learning and readies students for the use of performance-based assessments to judge learning. Many teachers use products as a way to "cover the content standards" in a more hands-on approach. Products can be developed during a lesson or as a culminating task and give teachers the opportunity to build stronger relationships with their students by acting as their hands-on learning facilitator. In essence, product work is an instructional method where students collaborate with others and "learn by doing." The same skills learned through this strategy are also many of the skills sought by employers. The World Economic Forum recently reported that coordinating with others and cognitive flexibility are two of the most important skills needed by students for the global economy (Schwab, 2018).

Core Strategy # 5: Targeted Thinking Skills. Strategically embedding thinking skills such as critical and creative thinking, problem solving and reasoning, and conceptual understanding through integrating macro themes into content prepares students for performance-based and real world assessment. In literature, for example, one might engage students in the following mental processes:

- Interpreting and creating analogies,
- Using deductive and inductive reasoning,
- Making inferences,
- Discerning authors' point of view or purpose, and
- Evaluating arguments.

In mathematics, thinking skills could be used for

- Looking for patterns, systems, and logical inferences based on algorithms,
- Guessing and testing,
- Making a model,
- Setting up an equation or applying an equation to an authentic problem, and
- Applying mathematical modeling, based on a series of assumptions.

In science thinking skills could be used to engage in scientific processes such as

- Selecting an issue for research,
- Making hypotheses,
- Collecting data,
- Analyzing and interpreting data,
- Drawing conclusions and testing their relevance to the problem,
- Making implications, and
- Replicating, as necessary.

As these core strategies suggest, implementing any or a combination of them into content areas will vary by discipline and by the nature of the required objectives. All of the core strategies are best taught within the context of content. In the prior teacher's case study, Ms. L. explicitly employed Core Strategies #2 through #5 prior to her culminating summative assessment.

Problem-Based Assessments

One type of performance-based assessment is problem-based. Gifted adolescents recognize that *problems* do not have one correct solution. Complex curriculum-based performance assessments will engage students in deliberate practice of exerting effort in both problem-finding and solution-finding. Problem-based learning (PBL), as an instructional strategy, is considered to be an innovative instructional method because of its emphasis on learning initiated by problems, self-directed learning, and collaborative learning in small groups (Koh, et.al., 2019). As a learner-centered approach, PBL provides complex, ill-structured problems that are rooted in real-life contexts and resemble situations that students are likely to encounter in a chosen profession (Stepien & Gallagher, 1993). Yet, using problem-based learning in assessment is much less researched and more controversial because assessment is probably the most important indicator for validating the effectiveness of PBL as an instructional approach (Savin-Baden, 2004). In their meta-analysis of PBL research, Gijbels et al. (2005) found that the effects of PBL varied, depending on the focus of the assessment instrument. PBL had the most positive effects when the instrument focused on assessing the understanding of principles that link concepts. In addition, students taught in a traditional manner performed better in basic knowledge acquisition while students taught with a PBL approach did better in application of knowledge and critical reasoning. An example of using a problem-based approach as an assessment task is illustrated in a teacher preparation program (for elementary education) from the University of Calgary (Koh et al., 2019):

> *I had the students build a rubric for a fire-drill, providing criteria for what a safe and effective fire drill might look like. We discussed the necessity of this task in an elementary classroom, and the value of introducing rubrics early in the year with students.*

The instructors reported that the value of coupling authentic problem-based assessments added to their teacher candidate's learning about assessment and real-world school problems. They shared that "student teachers were able to grasp the principles of authentic assessment in supporting assessment for and as learning practices when using a PBL approach" (Koh et al., 2019, p. 22).

Another example of a problem-based assessment in writing may be found in the Center for Gifted Education's William & Mary's language arts units (Center for Gifted Education, 1998). This final writing assignment has middle school students summarize and synthesize their unit learning on Romanticism in nineteenth-century America through a persuasive essay which was assessed with a persuasive writing rubric and both can be found through the Kendall Hunt website: https://k12.kendallhunt.com/program/william-mary-gifted-language-arts-curriculum

> You have been assigned the role of reporter for the library division of the *New York Times.* You have been given an assignment to synthesize the contributions of Romanticism in 19th Century America from the perspective of positive change that the period ushered in. You will need to consult various sources from the period, cite relevant literary examples, and make an argument for the positive role that the movement made. By the way, the piece can only be 750 words long, given the space allocated, and must be finished by 10:00 pm. tonight.

While PBL has been used more frequently as an instructional tool, teachers should consider the value in creating problem-based assessments where the emphasis is on the application of key concepts in problem solving and allow gifted students the opportunity to demonstrate their reasoning and understanding of content.

Product-Based Assessments

Product-based assessment is another form of performance-based assessment where products are developed by secondary students, typically at the culmination of a unit of study. Students have the opportunity to choose an area of interest or idea to explore in depth, analyze complex content, and create an authentic product, which either responds to specific requirements or is self-directed (Tomlinson & Moon, 2013; Renzulli & Reis, 2014). In the Schoolwide Enrichment Model (Renzulli & Reis, 2014), students completing a Type III activity, choose a real problem as the basis for conducting a long-term investigation, resulting in a unique student-driven product. The Type III product assessment requires assessing the goal and processes of the product development as well as the product itself.

Several curriculum models employed in gifted education involve product development as part of the instructional design and expected outcomes, rather than a paper-and-pencil test, so that students can integrate more complex thinking and demonstrate their learning in ways that represent appropriate expectations for gifted students involving advanced content, sophisticated processes, and authentic production. Product assessment usually is summative, but it can also be used to make inferences about next instructional steps. For example, gifted students often learn at a faster pace and can absorb and retain information more than their age peers, but even though the complexity of thought is evident, it may not come with organization skills, self-regulation skills or successful completion of a product. Therefore, incorporating processes that allow for monitoring progress and providing formative feedback should be considered. Further, teaching students to monitor and assess their own learning and progress toward product completion is critical in developing self-directed learners (Renzulli & Callahan, 2008).

The Role of Competition in Assessment

Another type of performance-based assessment that occurs outside of school but plays a key role in developing talent are competitions. Secondary students choose to be a part of a competition sometimes as a club, during or after school, or as part of an activities period. The role of competitions at the secondary level is key to promoting gifted adolescent performance because students can supplement, extend, and enhance the in-school experiences in a particular subject area or in a cross-disciplinary way. They are extracurricular options for gifted adolescents and regardless of the competition there are process and product outcomes. Typically, when a student is involved in a competition such as Science Olympiad or Model United Nations (U.N.), there are assumptions that the student self-selects, and the competition is in their talent domain, thus lending credence to the importance of the competition in supporting students' self-efficacy and personal goals (Olszewski-Kubilius, 2015). According to Olszewski-Kubilius (2015), competitions support the talent development stage of competency to expertise by providing opportunities to do authentic, creative work, connect with peers and other professionals and develop psychosocial skills and positive attitudes toward competitions and self-promotion. Competitions allow students to engage successfully with colleagues and mentors, take risks, experience self-confidence, and respond gracefully to critique (Subotnik & Jarvin, 2005).

The assessment for competitions is usually established by professionals in a discipline or field and represents a set of expectations that a professional in the field would have. There is a plethora of competitions for high school students ranging from academic, arts and music, writing and volunteering. Most extracurricular competitions result in a performance, product, or presentation by a student or a small team. They all have specifications and constraints related to the final assessment needed by the student or team of students, and many of them have state, national, and international levels of competition. Competitions may be the most authentic type of assessment for the gifted because they assess multiple higher level tasks linked to students' interests to real-world audiences at professional levels of expertise.

Challenges in Implementing Performance-Based Assessments

Performance-based assessments, including problem-based and product-based, afford secondary gifted students the opportunity to demonstrate growth over time. However, there are challenges in implementing such assessments.

Secondary schools have systems in place for grading, class rank, and preparing for benchmark assessments, which may impede the development and implementation of performance-based assessments. These existing systems can be a barrier and need some alteration to accommodate better the assessment of gifted learners. They should be supplemented, perhaps over time replaced, with more performance-based measures that provide better evidence of learning. While performance-based assessments can be lengthy to construct and evaluate, there are prototypes that have been developed to assist educators, aligned with content standards, which teachers can review and employ, as appropriate.

Another challenge with implementing performance-based assessments is using rubrics to evaluate specified products or problems. In outlining criteria to be used to ensure that standards are being met in highly specified tasks, Callahan (2002) recommends the following:

- Clearly delineate the dimensions (content, form, style) that the teacher will use in evaluating student performance and ensure that those dimensions reflect the instructional goals across content, process, and product dimensions.
- Find a way to express the highest level of quality on each criterion. Help students learn the difference between "best in class," "best that I can be," and "expertise."
- Help students understand the new assessment structure.
- Judge students by what they actually produce, not what you expect of them. Avoid "halo" effects that come from giving students the benefit of the doubt or rate effort rather than productivity, or negative assessment that comes from rating students lower because teachers "expected more." If more is expected, it should be specified in the rubric.

Kohn (2006) provides some cautions about overusing rubrics. He states that if students use them for every task they do, then too much attention is given to the quality of one's performance and may lead to more superficial thinking, less interest in whatever one is doing, less perseverance in the face of failure, and a tendency to attribute the outcome to innate ability and other factors thought to be beyond one's control. To that extent, more detailed and frequent evaluations of a student's accomplishments may be quite counterproductive.

While challenges exist to implementing performance-based assessments, the benefits appear to outweigh the barriers, especially given the difficulty of discerning gifted performance at secondary levels.

Role of Administration

School leaders play an important role in allowing and supporting the use of performance-based assessments schoolwide. If schools want to really know what students know and are able to do in order to raise expectations, they need to consider supporting and impeding structures for implementing and sustaining performance-based content assessments (see Table 9.1).

Administrators in gifted education, possibly more than in other fields, must be resilient because they are advocating for a group of students who display exceptional cognitive ability yet challenge the sensitivities of critics who contend that appropriately differentiated academic experiences for highly able children are somehow unfair to other children (Spielhagen & Brown, 2008). They must know when to push for change and from whom they can garner support. Supporting the use of performance-based assessments along with more traditional measures in schools require administrators to ensure that the multidimensional components of assessment are implemented effectively and coherently.

Conclusion

The assessment of gifted students requires a differentiated approach that combines traditional and performance-based measures in order to judge gifted student learning

TABLE 9.1 Supporting and impeding structures for implementing performance-based assessments

Components	Supporting	Impeding
Professional Development	Strong professional development program with sustained renewal opportunities. Strategic decisions are made about the professional development aligned with school improvement goals and mission statements.	Eclectic approach to professional development. For example, sending new hires for AP training with limited support, conversation, or follow-up. Alignment between assessment and content only exists within a particular course.
Fiscal Allocation and Support	Fiscal support for curriculum, materials, developing performance-based tasks ensures fidelity of implementation across content areas.	Fiscal support for only professional development but not necessary support materials; or support materials without professional development lacks coherency.
Leadership	School and central office provide strong leadership to support the concept of higher expectations for all students, through the use of performance-based and authentic assessments.	School and central office focus on remediation and those students below expected proficiency levels and efficient assessments that are easy to implement.
Accountability	Off level exams (including AP & IB) and other performance-based summative measures demonstrate accountability for learning linked to impact and connections to higher education.	Preparation for state testing has led administrators and teachers to focus on test preparation and interim assessments to the exclusion of other measures of assessment.
Access	Targeted and proactive measures are taken to ensure that a diverse body of students have access to performance-based assessments.	School policies, grading, class rank, and pre-requisites limit opportunities for ensuring all students have access to high end learning.

and make instructional choices. Use of off-level content-based assessment with gifted secondary students has yielded strong evidence of learning gains in specific curriculum areas through such vehicles as Advanced Placement and International Baccalaureate programs. Moreover, performance-based assessment approaches have also shown promise for demonstrating short term learning in areas such as scientific research skills, persuasive writing, and literary analysis (VanTassel-Baska & Little, 2023) and in enhancing critical thinking and reading comprehension (VanTassel-Baska et al., 2009). Product assessments too have provided important evidence of student growth in affective areas. These tools can supplement existing traditional measures to provide a comprehensive portrait of the learning impacts of our best students and presage the true potential they possess.

References

Abeywickrama, P. (2012). Rethinking traditional assessments concepts in classroom-based assessment. *The CATESOL Journal*, *23*(1), 205–213. https://files.eric.ed.gov/fulltext/EJ1112037.pdf

Abrams, L., M., Pedulla, J. J., & Madaus, G. F. (2003). Vies from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice*, *42*(1), 18–29. https://doi.org/10.1207/s15430421tip4201_4

Assouline, S., Colangelo, N., VanTassel-Baska, J. & Lupowski-Shoplik, A. (2015). *A Nation Empowered: Evidence Trumps the Excuses Holding Back America's Brightest Students*. The Connie Belin & Jacqueline N. Blank International Center for Gifted Education and Talent Development, College of Education, University of Iowa. Author. https://files.nwesd.org/website/Teaching_Learning/HiCap/2015-16%20meetings/NationEmpowered%20Vol1.pdf

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press

Callahan, C. M. (2002). The ABC's of creating a performance assessment task and scoring rubric. *Gifted Education Communicator*, *33*(2), 12–15.

Center for Gifted Education (1998). *Threads of change in 19th century American literature: A language arts unit for high ability learners*, grades 7-11. William & Mary. Kendall-Hunt.

Chappuis, J., & Stiggins, R. J. (2017). *An introduction to student-involved assessment for learning, 7e*. Pearson Education.

Chun, M. (2010). Taking teaching to (performance) task: Linking pedagogical and assessment practices. *Change: The Magazine of Higher Learning*, *42*(2), 22–29. https://eric.ed.gov/?id=EJ878687

Colangelo, N., Assouline, S., & Gross, M. (2004). *A nation deceived: How schools hold back America's brightest children* (Vol. 2). The University of Iowa, Connie Belin & Jacqueline N. Blank International Center for Gifted Education and Talent Development. https://www.accelerationinstitute.org/Nation_Deceived/ND_v2.pdf

College Board (2013). *AP: A foundation for success*. https://files.eric.ed.gov/fulltext/ED558096.pdf

College Board (2015). *AP courses*. https://apstudent.collegeboard.org/apcourse

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning*. Council of Chief State School Officers. https://edpolicy.stanford.edu/sites/default/files/publications/performance-counts-assessment-systems-support-high-quality-learning.pdf

Dixon, F. A., & Moon, S. M. (2015). T*he handbook of secondary gifted education* (2nd ed.). Prufrock Press.

Dougherty, C., Mellor, L., & Jian, S. (2006). The relationship between Advanced Placement and college graduation. National Center for Educational Accountability.

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Finnish Board of Education (2004). *National core curriculum for basic education 2004*. www.oph.fi/english/publications/2009/ national_core_curricula_for_basic_education

Foust, R. C., Hertberg-Davis, H. H., & Callahan, C. C. (2008). "Having it all" at sleep's expense: The forced choice of participants in Advanced Placement courses and International Baccalaureate programs. *Roeper Review*, *30*(2), 121–129. https://doi.org/10.1080/02783190801955293

Gijbels, D., Dochy, F., van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational* Research, 75(1), 27–61. https://doi.org/10.3102/00346543075001027

Hargrove, L., Godin, D., & Dodd, B. (2008). *College outcomes comparisons by AP and non-AP high school experiences* (College Board research report no. 2008-3). College Board. https://eric.ed.gov/?id=ED561030

Karpov, Y. V., & Bransford, J. D. (1995). Vygotsky and the doctrine of empirical and theoretical learning. *Educational Psychologist*, *30*(2), 61–66. https://doi.org/10.1207/s1532695ep3002_2

Keng, L., & Dodd, B. (2008). A comparison of college performance of AP and non-AP student groups in 10 subject areas (College Board report no. 2008-7). College Board. https://files.eric.ed.gov/fulltext/ED561028.pdf

Koh, K., Delanoy, N., Bene, R., Thomas, C., Danysk, G., Hone, G., Turner, J., & Chapman, O. (2019). The role of authentic assessment tasks in problem-based learning. *Papers on Postsecondary Learning and Teaching*, *3*(1), 17–24. https://files.eric.ed.gov/fulltext/EJ1302881.pdf

Kohn, A. (2006). The trouble with rubrics. *English Journal*, *95*(4), 12–15. https://eric.ed.gov/?id=EJ751576

Nash, G., Jeffrey, J., & Roy, J. (2009). *The American people: Creating a nation and a society*. Pearson.

Olszewski-Kubilius, P., & Thomson, D. (2014). Talent search. In J. A. Plucker and C. M. Callahan (Eds.), *Critical issues and practices in gifted education. What the research says* (2nd ed., pp. 633–644). Prufrock Press.

Olszewski-Kubilius, P. (2015). The role of out-of-school programs. In F. A. Dixon & S. M. Moon (Eds.), *The handbook of secondary gifted education* (2nd ed., pp. 261–282). Prufrock Press

Renzulli, J., & Callahan, C. (2008). Product assessment. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 259–284). Prufrock Press

Renzulli, J., & Reis, S. (2014). *The schoolwide enrichment model: A how-to guide for talent development* (3rd ed.). Prufrock Press

Savin-Baden, M. (2004). Understanding the impact of assessment on students in problem-based learning. *Innovation, Education & Teaching International*, *41*(2), 223–233. https://doi.org/10.1080/1470329042000208729

Schneider, J. (2009). Privilege, equity, and the Advanced Placement Program: Tug of war. *Journal of Curriculum Studies*, *6*, 813–831. https://doi.org/10.1080/00220270802713613

Schwab, K. (2018). *The global competitiveness report*. World Economic Forum. www3.weforum.org/docs/GCR2018/05FullReport/TheGlobalCompetitivenessReport2018.pdf

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066–1011). American Education Research Association.

Spielhagen, F., & Brown, E. (2008). *Excellence versus equity: Political forces in the education of gifted students*. www.routledgehandbooks.com/doi/10.4324/9780203887875.ch20

Stepien, W., & Gallagher, S. (1993). Problem-based learning: As authentic as it gets. *Educational Leadership*, 50(7), 25–28. https://eric.ed.gov/?id=EJ461126

Subotnik, R. F., & Jarvin, L. (2005). Beyond expertise: Conceptions of giftedness as great performance. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed.). Cambridge University Press.

Subotnik, R., Olszewski-Kubilius, P., & Worrell, F. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Suldo, S., M., Shaunessy-Dedrick, E., Ferron, J., & Dedrick, R. F. (2018). Predicator of success among high school students in advanced placement and international baccalaureate programs. *Gifted Child Quarterly*, *62*(4), 350–373. https://doi.org/10.1177/0016986218758443

Theokas, C., & Saaris, R. (2013). *Finding America's missing AP and IB students*. The Education Trust and Equal Opportunity Schools. https://edtrust.org/wp-content/uploads/2013/10/Missing_Students.pdf

Tomlinson, C., & Moon, T. (2013). *Assessment and student success in a differentiated classroom*. ASCD.

VanTassel-Baska, J. (2008). *Alternative assessments with gifted and talented students*. Prufrock Press.

VanTassel-Baska, J. (2013). Performance-based assessment: The road to authentic learning for the gifted. *Gifted Child Today*, *37*(1), 41–47. https://doi.org/10.1177/1076217513509618

VanTassel-Baska, J., & Baska, A. (2019). *Curriculum planning & instructional design for gifted learners* (3rd ed.). Prufrock Press.

VanTassel-Baska, J., Bracken, B., Feng, A., & Brown, E. (2009). A longitudinal study of enhancing critical thinking and reading comprehension in Title I classrooms. *Journal for the Education of the Gifted*, *33*(1), 7–37. https://doi.org/10.1177/016235320903300102

VanTassel-Baska, J., & Little, C. (2017). *Content-based curriculum for high ability learners* (3rd ed.). Prufrock Press.

VanTassel-Baska, J. & Little, C. Eds. (2023) *Content-based curriculum for gifted learners* (4th ed.). Routledge.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, *2*(2). https://doi.org/10.7275/ffb1-mm19

Wiley, A., Wyatt, J., & Camara, W. J. (2010). *The development of a multidimensional college readiness index*(College Board Research Report 2010-3). The College Board. http://professionals.collegeboard.com/profdownload/ pdf/10b_3110_CollegeReadiness_RR_WEB_110315.pdf

CHAPTER

10

Project-Based Assessments

Tasks and Rubrics

Catherine M. Brighton, Michelle Hock, and Tonya R. Moon

Assessment and instruction are inextricably linked. Using a backwards design framework, assessments should be crafted to assess the overarching learning goals for a unit, and instructional experiences should then be designed in order to support success on those assessments. For the purpose of this chapter, we therefore necessarily discuss both assessment and instruction as mutually reinforcing elements of classroom practice. Although we foreground the assessment component of performance tasks, we also discuss instructional considerations that must accompany the development and implementation of any performance assessment.

Project-based tasks and rubrics, including performance assessments, are alternatives to traditional types of assessments, and offer opportunities for all students, including advanced learners, to engage in the application of their learning instead of passive recitation of knowledge and skills (Miller & Krajcik, 2019). In classrooms employing project-based tasks, students investigate issues, problems, or opportunities embedded within local, national, or international contexts that have relevance to their interests and are appropriate for their developmental stage. Often these project-based assessments (PBAs) focus on problems that are meaningful to students because of the direct connection to their community, which may provide opportunities for mentorships with working professionals (Han & Bhattacharya, 2010). PBAs may entice students by utilizing technical skills (e.g., development of an app to solve a problem) and bolstering their communication skills.

"Non-traditional" assessments, like PBAs, have other potential benefits for student learning if designed and implemented well. These potential benefits include demonstrating a deeper, more connected level of knowledge (Schneider et al., 2016); increasing student motivation in both the learning and assessment contexts (Shin, 2018); developing students' critical, creative, and metacognitive skills (Surahman et al., 2018); and offering opportunities for increased access to post-secondary careers, particularly for students at the secondary level (e.g., LaForce et al., 2017). Despite these benefits for advanced learners, there are also potential liabilities, particularly surrounding issues of equity and access for under-represented groups of students and diverse classroom contexts that need attention when considering the implementation of PBAs. Moreover, there are technical adequacy concerns for both validity and interrater reliability for these types

DOI: 10.4324/9781003285991-15

of assessments. In this chapter, we focus on PBAs from a practical, not technical lens, including considerations for employing them.

Definitions and Characteristics

Although no single definition of PBAs exists, there are several features of these tasks that are commonly agreed upon. First, PBAs are in stark contrast to the traditional "teach and test" process. That is, to be a PBA, it has to occur over an extended period of time rather than as a culminating activity at the end of a series of lessons on a narrow topic. The implementation of PBAs requires teachers to serve more as facilitators of learning throughout instruction (i.e., coach more, instruct less), get out of "silo-thinking" by recognizing the multidisciplinary/interdisciplinary aspect of learning, and be comfortable with a degree of uncertainty and discovery for students as they engage in these types of assessments. PBAs require complex thinking and critical reasoning skills and call for students to evaluate, explain, and defend their ideas. PBAs also call for students to engage in the analysis of important and substantive problems for which there is no single correct answer. Rather, students may work towards creating an improvement in an important, defined space (e.g., local community).

Here's an example PBA in an elementary setting, where students are studying humans' impact on climate change. As part of the unit of study, students learn about carbon footprints and their personal roles in climate change. One relevant PBA for the unit might require students to investigate their own footprint (and/or their family's footprint) for the purpose of identifying ways to reduce it. It might also require students to provide recommendations to the larger community (e.g., school board, city's planning commission) to promote footprint-reduction. To do this successfully, students must develop deep understandings of the content (e.g., climate change, carbon footprints, mathematical concepts such as graphing) and the context of the task (e.g., their local community). They must then draw upon those understandings, as well as upon knowledge they acquire and skills they develop throughout the instructional cycle, to propose a solution to a problem, create a product, or demonstrate a performance. Generally, this requires students to tap into higher-order thinking skills such as making inferences and predictions, synthesizing information, making cross-disciplinary connections, and/or generating and testing hypotheses. A "non-example" of a PBA on this topic would include assessments in the forms of quizzes during instruction and the end of the chapter test on the last day of the unit. In this "non-example" all students take the same quizzes and tests which are scored with a common "answer key."

In contrast, project-based assessments employ a rubric, which is a scoring guide that explicitly describes graduated performance levels ranging from "novice" to "exceeding expectations." The rubric for the PBA described above might focus the assessment on the areas of mathematics (e.g., graphs), communication (e.g., writing, public speaking), evidence support, inferences, and proposed solution The example below in Table 10.1 highlights what the graphing domain of the overall rubric might look like.

A Comparison of Project-Based Learning to Problem-Based Learning

While project-based learning and problem-based learning have very similar-sounding names and share similar traits, they also may have key differences along the lines of size, scope, and end product.

TABLE 10.1 Graph domain from PBA carbon footprint rubric

Domain	Exceeds	Proficient	Needs Improvement	Not Acceptable
GRAPH	• Title and axes are accurately labeled; understanding of labels is well developed • Measurement units are identified on axes and contain proportional increments that encompass the data set • Precise plotting of all data points	• Title and axes are labeled, but minor misconceptions in the labeling may exist; understanding of labels adequately demonstrated • Measurement units may not be fully identified on axes and/or increments may not be proportional on one axis • Precise plotting of most data points; may be a few minor errors that do not substantially detract from graph quality	• Title and axes are labeled, but a major misconception in the labeling may exist; understandings of labels appear limited • Inconsistent and disproportionate measurement units across axes • Imprecise plotting of several data points; may be missing data plots in graph for one or both axes; may be several errors that detract from graph quality	• Title and axes are not labeled or several major misconceptions in the labeling may exist; understanding of labels not yet demonstrated • No measurement unit identified on axes; no consideration of proportionality of increments on axes. • Imprecise plotting of most/all data points; missing data plots in graph for both axes; numerous errors that substantially detract from graph quality

Similarities. Both approaches actively engage students in authentic tasks where the teacher serves as a facilitator of learning rather than as the deliverer of content knowledge. In addition, students typically work in small groups over an extended period, with formative assessments occurring throughout both approaches. A group product and presentation emerge as the bases for assessment of learning in both approaches.

Differences. As noted earlier, instruction and assessments are integrally linked, but the emphasis in problem-based learning is on instruction; the emphasis in project-based learning is assessment. Project-based learning begins with a clearly identified task, with an emphasis on the end product that is completely driven by content standards whereas problem-based learning starts with an ill-structured real world problem that students, as professional stakeholders (e.g., engineers, scientists), must address within ambiguous directions. They build on their prior knowledge about the problem and identify new knowledge that is needed. The emphasis in problem-based learning is on the process of acquiring learning rather than the end product.

(For a more thorough description of the use of problem-based learning scenarios with gifted learners and the evidence of meta-analytical learning gains, please see Chapter 9.)

Teaching and Learning through Engagement in Tasks (PROCESS rather than an EVENT)

Project-based assessments facilitate the interconnection of curriculum and assessment by providing a context for students to make sense of what they are learning (curriculum) by engaging in extended projects that, when completed, serve as the evidence of their learning (assessment). Students' engagement with these tasks and aligned rubrics require the application of knowledge, skills, and understandings in order to create a product or demonstrate a performance while simultaneously engaging in meaningful and relevant learning activities. In this way, PBAs stand in contrast to more traditional assessment methods, which often emphasize passive recitation of information or the use of lower-order thinking skills *after the learning process* that disconnects learning from assessment. Because PBAs call for the use of higher levels of thinking, they often occur over longer time periods and are embedded *within the learning process* as a part of the instructional cycle. Again, this contrasts with more commonly used one-time-event assessment methods, such as quizzes and tests. McTighe (Palmarini, 2011) refers to these as "cornerstone assessments" and, like the cornerstone of a building, these products or performances anchor key learning goals.

Identifying Learning Goals for Assessment

For PBAs to serve as effective assessment tools, they must be built upon clear learning goals that are communicated to students. Because PBAs are both learning and assessment opportunities, it is essential to keep sight of the importance of academic standards and the role that they play in these types of assessments; PBAs are not simply implemented as a culminating activity but rather for continuous learning throughout the instructional process. The following points should be considered when designing learning goals for PBAs so that they stay structured within the teaching process:

Not All Standards Are Created Equal

It is important to note that some standards focus on the acquisition of content knowledge/skills (ranging from low-level facts to high-level abstract ideas or skills); while other standards emphasize meaning-making (e.g., comparing and contrasting) or require the transfer of knowledge/skills in novel ways or into new settings. PBAs are based on academic standards that allow for the application and/or transfer of knowledge and skills in context rather than simply rote recitation, recognition, or decontextualized application. When determining PBA goals, the most important standards, or power standards, should be incorporated whenever possible. Examples might be in ELA, analyze multiple texts to compare theme; in math, evaluate different approaches to problem-solving; in science, create an experiment to test a question; in history, demonstrate that "events make a leader," providing evidence from multiple examples.

Standards Must Be "Unpacked"

Academic standards are a type of nonfiction text; therefore, one must analyze and interpret the meaning behind the standard within the context of its teaching. Dedicate time to talk through the meanings of the standards and identify the specific learning goals. What will each student *independently* know (K), understand (U), and be able to (D) demonstrate (KUDs) as a result of this PBA? Sometimes students work independently; however, it is more common that groups of students collaborate on a single PBA. Consequently, it becomes imperative that at the completion of the PBA, the teacher can speak to each student's individual level of achievement.

Seek to Design Multi-Faceted Learning Goals

Authentic engagement in a real-life scenario (or simulation) is an important feature of PBAs. Therefore, PBAs should be inclusive of the process as much as the final product or outcome. In addition to the content and skill goals across multiple disciplines, PBAs should be situated within realistic conditions and constraints that are often community-oriented. Goals may also include: (a) twenty-first-century goals; (b) Habits of Mind; (c) personal challenge goals; and (d) executive functioning goals.

Clarifying Assessment Criteria

Drawing on backwards design principles (Wiggins & McTighe, 2011) it is important to begin with the end in mind. However, for PBAs within real-world contexts (e.g., local community), the outcome cannot always be fully known. Even so, for a PBA to be an effective measure of student learning, clearly defined and observable evaluation criteria that are measurable should be identified and operationalized prior to implementation. For example, there should be evidence provided by the student in their response that they have demonstrated the identified criteria. In other words, using descriptors like "Demonstrates Understanding" or "Knows" is not easily (or reliably) measured; rather, what does a student "DO" that demonstrates understanding. Returning to the earlier rubric domain of "Graphs" (Table 10.1), one clearly sees what the student's response should contain (e.g., correct plotting of data points) as opposed to simply indicating "Has a graph."

Choosing power standards and paying attention to essential meanings of the project design reaps dividends. This first step in developing a rubric for a PBA involves identifying domains, based on the learning goals established for the PBA. Analytic rubrics are two-dimensional scoring guides that describe graduations of performance (e.g., Expert to Novice) across independent multiple domains (e.g., Graphing, Communication). The second step is to describe the performance indicators, or quality attributes, to support the PBA. Often, these indicators exist on a continuum, ranging from "exceeding" a criterion to "not yet meeting" a criterion. Using a well-designed, coherent set of criteria (i.e., rubric) that is inclusive of meaningfully different performance quality levels across a continuum allows for:

- a clear articulation of the expectations for student performance;
- a more efficient and consistent method for providing feedback (Rose/Thorn/Bud: what is good [rose], what is bad [thorn], and what is potential [bud]);
- communication to students of the expectations and critical issues in a PBA;
- shared expectations of learning and grading practices within a classroom, grade level, or school, resulting in the potential for more consistent, reliable grading; and
- self-evaluation or evaluation of peers' performances, an important component of learning.

In other words, if students demonstrate proficiency relative to the performance criteria, this should be an indication that they have met the learning goals being assessed.

In developing a rubric to assess a PBA in the area of Visual Arts, it is important to begin by describing the grade-level expectations (proficient). For example, the standards framework from the National Core Arts Standards (2014) in the area of

Visual Arts focuses on four areas: Observation and Learning to Comprehend; Envision and Critiquing to Reflect; Inventing and Discovering to Create; and Relating and Connecting to Transfer. At the lower elementary level, Connecting to Transfer proficiency, or grade-level expectations, might be defined as *being able to compare and contrast details in self-portraits from different times* (Table 10.2).

The next step in the development of performance criteria is to move *beyond expectations* to describe the characteristics, behaviors, or qualities that *exceed* each attribute/domain (Advanced/Exemplary) before then moving to the lowest level (e.g., Needs Substantial Work/Novice/Beginner). Continuing with our Visual Arts example, beyond expectations (Table 10.3) is defined as being able to *compare and contrast details in self-portraits and explain how specific details reveal information about the individual in the portrait*. Table 10.4 displays criteria for Needs Substantial Work, the next level defined, and Table 10.5 highlights the fully developed rubric in the area of Visual Arts Connections at the lower elementary level.

To ensure that all students get growth-focused feedback to facilitate their continued learning, it is important that teachers identify those aspects of the project where each student has done well in addition to what specific areas are either not yet evidenced or

TABLE 10.2 Step 1: Defining proficiency

	Advanced	Proficient	Limited Evidence	Needs Substantial Work
Connections		Compare and contrast details in self-portraits from different times providing only general basic supporting information.		

TABLE 10.3 Defining advanced

	Advanced	Proficient	Limited Evidence	Needs Substantial Work
Connections	Compares and contrasts details in self-portraits and explains how the specific details reveal information about the individual in the portrait.	Compares and contrasts details in self-portraits of one individual from different times providing only general basic supporting information.		

TABLE 10.4 Defining needs substantial work

	Advanced	Proficient	Limited Evidence	Needs Substantial Work
Connections	Compares and contrasts details in self-portraits from different times and explains how specific details reveal information about the individual in the portrait across time.	Compares and contrasts self-portraits of one individual from different times providing only general basic supporting information.		Compares and contrasts self-portraits but does not provide supporting information and/or may include irrelevant information.

TABLE 10.5 Complete connections rubric

	Advanced	Proficient	Limited Evidence	Needs Substantial Work
Connections	Compares and contrasts details in self-portraits from different times; explains how specific details reveal information about the individual in the portrait across time.	Compares and contrasts self-portraits of one individual from different times; provides only general, basic supporting information.	Compares and contrasts self-portraits of one individual; does not yet consider portraits from different times, providing limited supporting information.	Provides limited comparisons and contrasts of one self-portrait; does not yet provide supporting information that is relevant.

TABLE 10.6 Holistic rubric example

Exceeds	Proficient	Developing
Student expresses multiple interrelated ideas clearly and provides correctly formatted references and extensive within-text citations. Attention is given to the audience with no use of jargon or slang. Minimal grammatical or spelling errors such that meaning is not impeded.	Student's ideas are clear but there may be some inconsistent formatting of references and within-text citations. Attention given to the audience with occasional jargon or slang used. Some grammatical or spelling errors may occasionally result in minor impediments to meaning.	Student's ideas are disorganized, with few to no references used. No attention given to audience; jargon and/or slang is used throughout. Numerous grammatical or spelling errors occur that frequently result in major impediments to understanding.

require further attention. An analytic rubric separately describes the expected levels of performance for each of the components of the overall product to provide more targeted explanation of expectations (see Table 10.5). Analytic rubrics specifically describing performance gradations from "far below expectations" to "exceeds expectations" (in multiple domains) are helpful tools for teachers to guide targeted feedback for improvement.

At times there may be a desire to use a holistic rubric instead of an analytic one. A holistic rubric is a one-dimensional rubric that contains only a description of overall performance levels (i.e., not in specific domains). See Table 10.6 for an example of a holistic rubric for an oral presentation.

To maximize the benefits of PBAs, students should be engaged from the beginning. They should know the expected levels of performance so that they have transparent and meaningful targets for their efforts. Sometimes these criteria may be established and shared with students at the beginning of the project launch. In other cases, the teacher and students may work collaboratively to identify what the levels of performance would look like according to the context of each project. In both scenarios, it is critical that students fully understand what the evaluation criteria are, and how they are used to assess performance so that they have multiple opportunities to clarify the expectations well before assessment decisions are made. This is especially important for the successful implementation and completion of PBAs because students will likely be making performance decisions "in the moment." Having clearly-established performance criteria at the beginning of a project aids students in understanding what is expected. Criteria should also be revisited as needed throughout the instructional cycle so that students are reminded of expectations and so that they can continually self-assess their work against the given criteria.

Having a clear picture of the criteria used to evaluate students' work is also critical for teachers. Clear criteria serve as reference points when giving feedback to students as they work to complete the tasks and can therefore help teachers ensure that their coaching feedback is effectively guiding students towards successful task completion. Clear criteria are also important for teachers at the scoring phase. In order to evaluate students' completed products or performances reliably, teachers should use the criteria to determine the degree to which students met the identified learning goals represented in the criteria (Johnsen, 1996).

Ensuring Project Success (Choices, Groups, Scaffolds, Feedback)

Carrying out PBAs requires that students engage in iterative cycles of understanding, questioning, researching, applying through reasoning, developing and testing hypotheses, evaluating evidence, synthesizing evidence, and integrating teacher, peer, and self-feedback (Mergendoller et al., 2006). In other words, students are both learning and demonstrating their learning throughout the implementation of a PBA. Therefore, it is important to the success of these iterative cycles that students fully engage in the work and that teachers have pre-determined assessment points throughout the cycles. Ways to increase students' engagement in a PBA are to ensure that they have a voice in the implementation process, perhaps by allowing students to make choices at some decision points (e.g., what particular issue will be the focus; the modality of final response—PowerPoint, Community Letter, etc.), by having students participate in groups with scaffolds and supports (e.g., organizational/management structures), and by providing students with timely, focused, and intentional feedback to foster their continued growth towards their fullest potential.

Choices and Autonomy

Choice can be a powerful motivator, and PBAs should be designed to allow for choice-infused, interest-based differentiation. Students can be given choices around the issue(s) that they explore within a content area(s) or around the modalities in which they investigate the content or demonstrate their learning (e.g., podcast or video presentation). For example, in the climate-change PBA discussed earlier, teachers might offer interest-based differentiation by allowing students to choose which group's carbon footprint they want to examine (e.g., their family's, their school's, their local community's, etc.). Students might also be given choice when deciding how they want to present their recommendations for footprint reduction. Some students might opt to create a public-service announcement video while others may choose to develop a podcast. However, there is one important caveat: Teachers must be able to accommodate these choices by having the necessary resources in place to ensure that students can pursue their interests. Two important considerations teachers must account for when incorporating students' choice involve (a) ensuring that the learning outcomes remain common for all students and that, for whatever choice students make, (b) there are sufficient resources to allow for success. That is, students' choices—whether they involve choices related to methods for conducting their inquiry or involve different modalities for demonstrating their learning—should not alter the standards that students are required to demonstrate, with students meeting the same learning outcomes regardless of the choices made.

Project-based assessments also serve as tools for building students' voices and cultivating their abilities to function as increasingly autonomous learners. Because PBAs often incorporate student choice, engaging in these tasks may lead students to take greater ownership of their learning and to view themselves as drivers of their own learning processes (Stefanou et al., 2013).

Scaffolds

Because these assessments take place over extended periods of time, students need to monitor their own work and determine where they stand in relation to demonstrating the targeted learning outcomes. This provides students the chance to reflect on their progress and engage in self-directed learning, which also leads to an increased sense of autonomy. It may also require that additional management considerations be incorporated into the implementation plan to ensure that students continue to meet expected timelines for completion. Some advanced learners excel in these necessary executive functioning skills, while others, including some advanced learners, require scaffolds and supports as they work to demonstrate higher-order thinking and skill application. Providing these scaffolds involves pre-planning and structuring by the teacher, including considerations relating to time and resources needed for students to complete their work and demonstrate their learning. Asking students to reflect periodically on their progress in written form may also serve as a scaffold for forward motion.

Groups

PBAs are often designed as collaborative endeavors among students. Because of this, teachers should consider purposeful grouping strategies that are appropriate for the required tasks. These groupings may be based upon shared interests in a topic but may also allow for strategically assembling students with specific needs and talents. In addition to the content and skills of the PBA, the learning experiences should provide an opportunity to practice and develop students' collaboration skills, which may require students to engage in norm-setting or to adopt specific roles for task completion. An important consideration when grouping is that each student is responsible for demonstrating independently the identified learning goals. In order to verify every student's learning, there must be individual accountability for the learning goals. This does not preclude working in a team configuration, as being able to work effectively in a team is cited by the National Association of Colleges and Employers (2021) as a critical twenty-first-century career readiness indicator.

As critical as group functioning is, the grade for the PBA should be drawn from the individual student's achievement of identified learning goals. When planning for group work, teachers must design the tasks so that all students are *individually accountable* for meeting the identified learning outcomes. Teachers must therefore take this requirement into account when designing the PBA and make provisions to ensure that group work permits students to individually showcase their mastery of the learning outcomes.

Feedback

Just as in other "real-world" projects, successful PBAs involve an ongoing process of evaluation and revision—some formal and some informal. Likewise, teachers should plan dedicated check-in opportunities with individuals and/or groups of students as well as plan for informal check-ins as needed. The formal check-ins (e.g., individual conferences; submission of draft milestones) serve as a measure of incremental accountability toward pre-determined milestones and as opportunities to offer students further directions and feedback as needed. The informal check-ins, either scheduled or conducted "on the fly," provide a glimpse into group progress as well as a chance to listen in on their work in progress. A third approach may include written responses from students about their progress in the form of exit tickets (in person), individual conferences with students (in person or virtual) or through email surveys (virtual) to gauge their progress. Regardless of the forms "check-ins" take, focused feedback from the teacher should be ongoing throughout instruction as a necessary strategy to ensure successful task completion. This focused attention on student progress provides teachers with information on where students are in relation to the learning goals, how to best support their learning and adjust timelines as needed, and how to provide meaningful feedback for specific ways to improve. Because PBAs engage students in an in-depth exploration of a topic, continual focused feedback helps students reflect on their learning in relation to the learning goals and expectations and to identify actions that they can take in order to meet expectations. At various stages of the PBAs, this feedback can also come from peers or from the students' self-reflections as part of metacognitive development.

Connecting Real World/Authenticity to Students' Lives (Current Issues)

One of the primary advantages of using PBAs is the ability to make learning authentic to students' lives. Project-based assessments are, by design, focused on real-world issues. When appropriate and feasible, this also opens a window for students to gain mentorship experiences by working with adults in their communities or for younger students to work alongside older students in their school or district. Seeing the direct relevance to "real people" makes the process of working to address local problems more meaningful and purposeful because students see the direct connection to themselves, their families, and their communities. According to Parker and Igielnik (2020), children and young adults in the current school-age demographic are strongly motivated by work with connections to equity, social justice, and advocacy. Designing PBAs with issues of equity and social justice (e.g., topics related to civil rights, environmental toxins, etc.) in mind further increases the likelihood of students' persistence with the tasks. Further, as these issues are often complex and can be viewed differently from multiple stakeholder perspectives, this community-oriented advocacy work will also likely require that students demonstrate skills in questioning, researching, leading, acting, presenting, and reflecting by confronting multiple realities. Because students will be primed to engage in this work, it provides teachers with a rich opportunity to interject content (e.g., historical examination of other forms of oppression) and skills (e.g., action-research and presentation tools) used by community activists to help prepare them for this PBA work.

Connecting Across Disciplines

Project-based assessments have the advantage of being cross-disciplinary in nature, giving students the opportunity to draw upon or build their learning across a variety of content areas. The tasks involved in completing PBAs often require students to employ cross-disciplinary skills sets, such as the use of critical reasoning, argumentation, and inference-making. Because these skills cut across disciplines, PBAs require students to synthesize their learning from different content areas, to approach their work with varied frames of references, and to employ disciplinary skills in one content area that they may have developed in another area. In doing this, students begin to approximate the real-life work of problem-solving, which involves compiling ideas, resources, and skills from multiple disciplines. By nature, then, PBAs require students to adopt varied vantage points, which builds their perspective-taking skills and hones their problem-solving abilities.

Flexibility in Design: Online and Hybrid Formats

Even before the COVID-19 global pandemic changed the ways in which schools function, contemporary K-12 school settings have explored ways to better incorporate technology-enhanced learning opportunities for students. Another benefit of PBAs is the great flexibility they afford in terms of the setting in which they can be implemented. PBAs can easily be carried out in person, online, or in a hybrid delivery model allowing learning to occur beyond the finite time and space of a physical classroom session. Further, utilizing PBAs in a pull-out program or through an extra-curricular context provides even more flexibility for students and programs. The same principles for designing, implementing, and evaluating PBAs are applicable, regardless of the class context or delivery format. In any environment, the same non-negotiables apply. Teachers must ensure that there are clear learning targets that give students ongoing focused feedback about their progress in relation to the learning outcomes, and provide students with continued scaffolding, supports, and check-ins.

One advantage of carrying out PBAs in an *online or hybrid format*, however, is the ability to use collaboration apps that can support students' learning and can provide teachers with valuable assessment data. Although these apps can be integrated into in-person instruction, they may prove particularly useful for orchestrating effective group interactions when students work remotely or on different schedules throughout a day or week (based on class schedules). These collaborative apps can also serve as assessment tools that help teachers monitor students' collaborative efforts as well as the individual contributions of each group member. Some applications that are particularly useful for real-time video-conferencing and include screen-sharing capabilities are Zoom, Google Meet, Microsoft Teams and Canvas, which are often easy to access through free or school-based accounts. When using these applications, it is incumbent to ensure that they are approved by the district's central office regarding protection of student privacy. Because it is not always possible to gather groups together in real time, it is helpful to utilize tools that facilitate asynchronous idea generation such as Jamboard, Mural, Schoology, and Slack. It is not always strategic or even possible to deliver coaching feedback in person, and so drawing upon some effective tools for providing targeted feedback and coaching such as Voice Thread, Hypothesis, FlipGrid, and the SpeedGrader function within Canvas will ease the process of delivering timely virtual feedback. Ensuring organization over a sustained period of time benefits from project management tools such as Project Pals, Google Classroom, and the use of Trello Boards to keep tasks and timelines on every student's mind. Because technology is ever-changing, this list is just a starting place for teachers to find the tools and structures that work best in their contexts when assessing students' learning through PBAs.

Considerations for Implementation

In the previous sections, we outlined the rationale for and benefits of utilizing PBAs to meet the needs of advanced learners in a variety of educational contexts. As with any educational effort, success depends on careful advanced planning. Therefore, in the following section, we outline several suggested implementation strategies to optimize success.

Pre-Implementation Planning

1. *Ensure an appropriate project scope*. Because PBAs are part of the instructional process (i.e., woven throughout instruction for the duration of a unit) and therefore do not occur in the space of a single day or class period, teachers should be mindful of the scope of the projects students undertake. Given the rich potential to expand and connect to other content areas and real-world problems, it is easy to imagine how the scope of the projects could quickly exceed available time. For this reason, it is critical that teachers realistically consider what space is available within the learning sequence for this purpose. Knowing that the creation of products and performances may take several days (or even weeks) and that students need to receive feedback and continually make revisions to their work, teachers must plan a practical schedule for completion and ensure that the scope is manageable, even mapping specific instructional blocks for this work. Time allocation will vary across grade levels and the classroom context (e.g., if this work is occurring within a general education classroom or part of a pull-out or enrichment class).
2. *Ensure adequate resources*. Resources are another logistical consideration that teachers must consider when planning for PBAs. Depending on the nature of the PBA, classroom resources (e.g., books, supplies) or technological tools (e.g., high-speed internet access, specific hardware or software) may be needed. Students may also need access to experts or professionals to support their work. Because ease-of-access for these resources cannot be assumed, teachers must carefully plan how they will secure needed materials or ensure that students can work with the appropriate adults. In the event that accessing resources reliably presents a barrier for students, teachers may need to reconsider elements of the PBA and find alternative means by which to support students' learning. In order for PBAs to function optimally and equitably, it is essential that teachers ensure that all students have appropriate access to material and human resources.

Implementing PBAs with Success: Virtual Book Clubs

After talking with the head librarian from the community library, Mr. Williams, the head librarian at a middle school, becomes concerned about the lack of variety in the books that students check out from the library. Mr. Williams decides to contact Ms. Gaston, an English Language Arts (ELA) teacher, to see if the students in her classroom similarly show little interest in a variety of genres. Ms. Gaston tells Mr. Williams that her students often confine their reading selections to only one or two genres. Mr. Williams, Ms. Gaston, and the community head librarian decide that it would be beneficial for students, the school community, and the larger surrounding community if they worked together to increase students' interests in a variety of book genres, so they brainstorm some ideas for how they can address this shared problem. They ultimately decide that Ms. Gaston, in collaboration with the gifted resource teacher, Ms. Shreve, could use a PBA to help students explore multiple literary genres (Table 10.7). In addition, they want to incorporate 21st-century skills into the PBA. Ms. Gaston's and Ms. Shreve's first step is to identify relevant power standards that they can unpack in order to ensure that they target higher levels of thinking. After this, they also generate a list of learning goals

TABLE 10.7 Sample project task and rubric

Example PBA

Middle School Library Virtual Literary Program
This PBA is designed to assess the depth of student understanding of various forms of literature and students' ability to communicate that understanding in a manner appropriate to a particular audience.

Power Standards
- Common Core Reading Standard 9: Compare and contrast texts in different forms or genres (e.g., stories and poems; historical novels and fantasy stories) in terms of their approaches to similar themes and topics. Includes the subgenres of adventure stories, historical fiction, mysteries, myths, science fiction, realistic fiction, allegories, parodies, satire, and graphic novels.
- Common Core Writing Standard 9: Draw evidence from literary or informational texts to support analysis, reflection, and research.

Learning Goals: Knowledge, Skills, and Dispositions

Students will demonstrate their ability to:
- apply understanding of the characteristics of various literary genres
- compare, contrast, and analyze relationships among literary genres
- read and understand information from varied sources
- support ideas with examples
- organize information in a powerful and effective manner
- communicate ideas appropriate to a specific audience

Students will demonstrate 21st-Century Skills:
- 4 C's (critical thinking, creativity, collaboration, communication)
- 5 Life Skills (flexibility, leadership, initiative, productivity, social skills)

Prerequisite knowledge/experiences:
- knowledge of literary genres and their unique characteristics
- experience analyzing literary forms and styles
- experience with communicating for a specific audience

PBA Prompt: Local Literary Book Club
Your school's head librarian recently contacted you and your peers to help with a campaign to increase student involvement in the library's virtual book club. In the librarian's annual report, he found that book club participation has been declining and that some literary genres were not being accessed and read as widely as others. As a result, the librarian has tasked you with two things: (1) find a way to boost the book club's enrollment by generating more interest in literature and (2) increase students' interests in a variety of literary genres.

To get you started on this task, the librarian has provided you and your group members with some background information about the book club. The club meets virtually every week for several hours to discuss a common book of students' choosing. Most often, the students want books that they can connect to personally. However, they have typically limited themselves to choosing the same genres every month, which concerns the librarian because she doesn't think that they have a full appreciation for the written word or for the many forms of literature that exist.

You will be evaluated on your depth of insight in understanding each genre for its unique characteristics, the quality of the details and examples you use to make your comparisons, the appropriateness of your communication for your specific audience, and the quality and appeal of your abstract. (See the rubric for specific criteria.)

(*continued*)

TABLE 10.7 Cont.

Project Rubric				
	Exceeds Expectations	**Meets Expectations**	**Almost There**	**Needs Major Work**
Depth	• Provides a sophisticated comparison of genres, evaluating each type in detail and identifying areas of similarities and differences among types • Analyzes the principles of literary genre using many substantive details	• Provides a comparison of genres with an evaluation of each type; identifies areas of similarities and differences, but further details would make the comparison more robust • Analyzes the principles of literary genre, but could include more details	• Provides a limited evaluation and comparison of genres, focusing mostly on similarities or differences (perhaps not both); absent details result in a superficial comparison • Provides an incomplete analysis of principles of literary genres	• Does not yet provide a comparison or evaluation of genres, identifying few/no similarities or differences; genres discussed only in isolation (not comparatively) • Provides only a summary of information (no analysis)
Comparisons	• Gives many vivid details and examples that represent the genres • Reinforces the similarities and differences among the genres so that the audience sees a clear rationale for their use • Makes many appropriate and illuminating comparisons to help the audience distinguish among genres.	• Gives some details and examples that represent the genres • Reinforces the similarities and differences among the genres, although the audience may not see a clear rationale for their use • Makes comparisons so that the audience can distinguish among genres; comparisons could be more informative such that audience's understanding is more likely	• Uses details and examples, but they are not illustrative of the genres • Attempts to reinforce the similarities and differences among genres; contains minor misconceptions • Attempts to make comparisons, but they are not yet fully effective in helping audience distinguish among genres; comparisons could be more familiar to/suitable for the audience	• Does not use details and examples to represent the genres • Does not make clear comparisons among genres on themes and character development. • Does not make comparisons or comparisons are largely ineffective in helping the audience distinguish among genres; comparisons need to be altered entirely so that they are familiar to and suitable for the audience

TABLE 10.7 Cont.

Project Rubric

	Exceeds Expectations	**Meets Expectations**	**Almost There**	**Needs Major Work**
Communication	• Product is tailored to the audience, showing a thorough understanding of the audience's familiarity (no assumptions made) with genres and an awareness of the product's purpose • Ideas flow in a logical and sophisticated way • Uses context-appropriate language	• Product is tailored to the audience, with only minimal assumptions about the audience's familiarity with genres such that the product's purpose is mostly clear • Ideas flow fluidly, but there is uneven development of ideas • Most language is context appropriate, but a few minor revisions could be made	• Attempts to tailor the product to the audience, although several ideas are likely unfamiliar, and the product's purpose is not entirely clear . • Inconsistent idea development. • Attempts to use context-appropriate language, but many minor or a few major revisions are needed	• Product is not tailored to the audience. Assumptions are made about what the audience knows, resulting in confusion • Ideas do not flow coherently • Language is not context appropriate

that they want their students to meet during the course of the PBA and that are aligned to the power standards and their school curriculum. Once Ms. Gaston and Ms. Shreve establish the standards and goals that frame their PBA, they work together to develop the full scope of the assessment, including the nature of the tasks that students will perform, the materials and resources that will be needed, the schedule during which the PBA will be completed, and the performance criteria that will be used to evaluate students' work. They also discuss equity-related concerns and establish plans for supporting all learners as they work to attain the learning goals. Their final plan for implementation involves alternating independent student-driven "project time" with other ELA learning activities on an A/B schedule. The following scenario describes how Ms. Gaston and Ms. Shreve implement the PBA within the 9-week grading period.

Step 1: Set Students' Expectations. Clear articulation of learning expectations, student behavior, and work ethics throughout the entire process should frame the launching of a PBA. Ms. Gaston introduces the idea of virtual book clubs and initiates a discussion about how students will undertake the work. The task includes specific responsibilities for achieving the goals. In the launch, the teacher reviews the expected timelines and the details of each step that students engage in. Ms. Shreve invites students to keep a journal about the genres of books they're reading and to post daily to an online log through the class learning management system (LMS) as a form of accountability for their sustained engagement through the assessment. She also reminds them to consult the planning sheet that accompanied the PBA to help support their planning.

Step 2: Form Student Groups. Most PBAs involve student collaboration, and it is important for a teacher to remember that student cooperation and teamwork are essential. Based on Ms. Gaston's knowledge of each student's strengths and growth points, she forms teams comprised of a diverse range of levels of content expertise and skills and makes clear that while students will be working together, their work will be individually assessed. She reminds students that each member is a valuable team member, and that there is the expectation that each member contributes to the project success. She summons up several positive examples from the last PBA where students' work was enhanced by collaboration and cooperation.

Step 3: Create the Plan. Ms. Gaston crafted the context of the project but invites students to be active participants in the construction of their specific response to the PBA. Within the project parameters, she invites students to work with their group members to develop their ideas, reminding them how critical a solid planning process step is in having a successful PBA. This process involves brainstorming ideas of where the project can go, identifying what the end result will be (or having a vision for it), and then mapping out or storyboarding how the team will get to the end result. Ms. Gaston circulates among the groups as they storyboard their plans and turn to building their knowledge and developing the project focus; Ms. Shreve helps connect teams with relevant "experts" such as bookstore owners, authors, and the community librarian and arranges virtual meetings with students on a rotating basis. The team-based work occurs concurrently with ongoing instruction, individual students' own research, and interacting with experts and community members relevant to their projects. Ms. Gaston provides scaffolds such as graphic organizers to help students organize their learning. She works to ensure that all students have roadmaps for achieving the expected outcomes, including a management plan and timeline for how they will achieve their vision. Ms. Gaston, with appropriate input from students, reviews the rubric with teams to ensure that the plans are well aligned to the evaluation criteria.

Step 4: Implement the Plan. Ms. Gaston and Ms. Shreve keep careful notes about their meetings with students and review daily entries that are uploaded to the LMS. The systematic use of these formative assessments help ensure that students are well-positioned to achieve the learning targets. The teachers closely monitor who is doing what work based on each student's relative strengths and growth points and frequently refer to the project rubric to ensure that students are working toward meeting project expectations. It is during this step that targeted feedback becomes imperative to ensuring that students stay on track and are in alignment with expectations. The teachers continue to implement formal and informal check-ins, especially as the groups initially launch into the work, to ensure that students' plans are clear and that the project goals (and the rubric) serve as their compass.

Step 5: Differentiate for Students' Successes. On alternating days from "project time," Ms. Gaston designs and implements ELA lessons that support the learning goals assessed in the PBA. Using the formative assessment data such as exit tickets, reviews of students' daily logs, and her notes from individual student conferences, Ms. Gaston attends to the needs of individual students and the class as a whole, making adjustments to the pace, resources, and scaffolds as needed. Teachers should use the formative data that they gather as part of the formal and informal check-ins and surveys to adjust conditions in response to their students' needs. This may include the suggestion of a daily timesheet to ensure that students have accountability for their assignments, or the adjustment of task conditions to allow for more student input. Using the rubric for the PBA as a compass, Ms. Gaston continues to provide targeted feedback relative to the performance criteria so that students see how they are progressing toward their intended learning outcomes. Embedding formative assessments within the instructional phase as well as during the PBA group activities will provide teachers with the information that they need to monitor students' work and to determine whether or not students' tasks are ready for the summative assessment.

Step 6: Present the Project Outcomes. Because one of the greatest values of PBAs is the value of the end result to someone else (e.g., community members), finding an appropriate outlet for dissemination is important. The Virtual Book Club task left open the possibility of many forms the final project could take, and Ms. Gaston and Ms. Shreve probed students about their plans for the final product in their frequent check-ins. Some students found that an oral presentation to the librarian was the most effective outlet for disseminating their project outcomes. Another group developed a website with helpful infographics to highlight the variety of genres the library offers. A third group created a public service announcement to educate other middle school students about the breadth of literary genres. A fourth group recorded and edited a series of interviews with authors about their books (representing multiple genres) to entice readers to consider checking the books out of the library. For Ms. Gaston and Ms. Shreve, the delivery of these presentations occurred in real-time through in-person channels; however, they might also have been recorded and submitted virtually. The additional benefit of having a written or recorded product is that students can re-read (or re-watch) and self-reflect on their contributions to the issue at hand.

Step 7: Assess the Final Project and Process. Students may consider that the authentic publication or presentation is the final step of the PBA process, overlooking the importance of evaluating the product. Ms. Gaston and Ms. Shreve, however, require that students take time to reflect on the project experience and complete the rubric from their own student perspectives. Ms. Gaston notes that when the students have the opportunity to contribute to the discussion about their work, they are more reflective about areas where they fully achieved learning targets and about those areas where additional evidence of mastery would be needed. If students have kept their focus on the pre-planned learning goals and evaluation criteria outlined in the rubric, they can use these tools as levers for self-reflection about their engagement in the work. Additionally, using the project rubric to guide the final assessment of the end product or performance will help minimize subjectivity in the grading process. It is also important for supporting metacognitive skills and self-regulation skills that students contribute self-evaluations, as well as peer evaluations, of the end results.

Reflecting on Equity-Focused Project-Based Assessments

Equity and inclusion are important considerations for the implementation of PBA in a classroom. Today's classrooms are a mosaic of students from different cultures, having different talents, backgrounds, orientations, and classroom expectations and experiences. Equity means that all students, regardless of their personal and social characteristics, should have the opportunity to solve complex projects that are locally (or globally) grounded. Berger (2003) writes that once students see themselves as capable, they are

TABLE 10.8 PBA self-reflection tool for equity

Area: My PBA	Reflection
…is set within an **authentic context** where students transfer school learning for addressing larger (often community) issues	
…is based on **power standards** that require students to think in complex ways about important content typically from multiple disciplines	
…is **student-centered** *Student Agency*: allows for engagement in tasks that are meaningful and relevant, often driven by student interests, giving students choice *Scaffolded*: allows for building student success and self-efficacy through the provision of supports throughout the process to aid in the mastery of learning	
…fosters **team collaboration** *Intentional Formation of Groups*: equitable learning environments are important and are partly achieved through the intentional formation of diverse groups of students representing various gender identities, races/ethnicities, cultures, languages, and abilities *Role Rotations* within groups to avoid stereotyping various demographics (e.g., gender/racial) so that students learn new skills and see themselves in new ways not limited by roles (e.g., gender roles, racial roles, etc.) *Reflect and Celebrate Diversity* through self- and peer-assessment, identifying the contributions of each other and of self	
…uses **formative assessment** for the purpose of revision and improvement (not grading), which creates a safe space where failure is accepted and valued for its learning power as it provides opportunity to analyze, rethink, and problem-solve	
…uses **summative assessment** and **reflection** for students to demonstrate their individual level of obtain of the learning targets as well as their overall contribution to the end result.	

never quite the same, and that in order for students to see themselves as capable of excellence, classrooms should present intellectual challenges set within authentic contexts requiring collaboration and reflection (Becton-Consuegra, 2020). The following self-reflection tool (Table 10.8) allows for teachers to consider the degree to which their PBAs support equity and access so that every project—and every student—can be successful.

Recommendations for Directors of Gifted Programs

Getting started using PBAs as part of a gifted program may at first present challenges for some teachers because it requires them to be first, and foremost, facilitators of learning through coaching students rather than providing direct instruction. Common challenges surfaced at first with PBAs and some suggestions for overcoming the challenges are listed below.

1. *Student choice*. It is important to acknowledge that in typical school settings, students often have no choice or control over their own learning. In a PBA classroom, it is important for students to have choice; therefore, getting to know students' interests and preferences is an important aspect of successful PBAs.

2. *Management of time and resources*. Because of the passive nature of school for most students, first implementing PBAs may result in students' not knowing how to manage their time, their freedom, or the unstructured nature of such assessments Therefore, it is important to set up PBAs so that students have the greatest opportunity to be successful. Such structure can come through tools students are provided to assist in time management and organization (e.g., templates, checklists, timelines, and deliberate, focused, and detailed feedback).
3. *Quality Products/Performances*. It is important to establish a culture that supports high-quality student work. Such a culture is established by holding high expectations for student work, providing consistent and meaningful feedback to facilitate student learning based on regular check-ins, allowing for students' trial and error to learn from missteps, and engaging students in self- and peer-reflections.
4. *Community support*. Because of the importance of students seeing the relevance of school learning outside of school, having strong community ties is especially important in successful PBAs. Community connections provide mentoring opportunities for students and access to culturally specific resources whereby students can see themselves in the learning process.

Summary

The driving question around project-based assessment that most educators want to know is "Does it have an impact on student learning?" While not exhaustive, studies (e.g., Almulla, 2020; Barshay, 2021; Duke et al., 2020; Saavedra et al., 2021) are beginning to add promising evidence that the adoption of a project-based teaching and learning approach does improve student learning in core content areas through the secondary level. Programs supporting students in disciplined project-based tasks allow advanced learners to participate authentically in meaningful learning experiences within a dynamic environment rather than simply being passive learners.

References

Almulla, M. A. (2020). The effectiveness of project-based learning (PBL) approach as a way to engage students in learning. *SAGE Open*, 1–15. https://doi.org/10.1177/2158244020938702.

Barshay, J. (2021). *Proof points: Four new studies bolster the case for project-based learning*. The Hechinger Report.

Becton-Consuegra, D. (2020). *High-quality PBL as an equity imperative*. www.pblworks.org.

Berger, R. (2003). *An ethic of excellence. Building a culture of craftsmanship with students*. Heinemann.

Bland, L. M., & Gareis, C. R. (2018). Performance assessments: A review of definitions, quality characteristics, and outcomes associated with their use in K-12 schools. *Teacher Educators' Journal*, *11*, 52–69.

Duke, N. K., Halvorsen, A-L., Strachan, S. L., Kim, J., & Konstantopoulos, S. (June 2020). *Putting PjBL to the test: The impact of project-based learning on second graders' social studies and literacy learning and motivation in low-SES school settings. American Educational Research Journal*, *58*(1), 160–200. doi.org/10.3102/0002831220929638

Han, S., & Bhattacharya, K. (2010). Constructionism, learning by design, and project-based learning. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology* (pp. 127–141). Jacobs Foundation.

Johnsen, S. (1996). What are alternative assessments? *Gifted Child Today*, *19*(4), 12–50. https://doi.org/10.1177/107621759601900407

LaForce, M., Noble, E., & Blackwell, C. (2017). Problem-based learning (PBL) and student interest in STEM careers: The roles of motivation and ability beliefs. *Education Sciences*, 7(4), 92. https://doi.org/10.3390/educsci7040092

Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2006). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. *The Interdisciplinary Journal of Problem-Based Learning, 1*(2). https://doi.org/10.7771/1541-5015.1026

Miller, E. C., & Krajcik, J. S. (2019). Promoting deep learning through project-based learning: A design problem. *Disciplinary and Interdisciplinary Science Education Research*, *1*. https://doi.org/10.1186/s4303101900096

National Association of Colleges and Employers. (2021). *The attributes employers seek on students' resumes. Job Outlook 2021 Spring Update*. www.naceweb.org/talent-acquisition/candidate-selection/the-attributes-employers-seek-on-students-resumes/

National Core Arts Standards. (2014). *National Core Arts Standards: Conceptual Framework*. Retrieved from www.nationalartsstandards.org/content/national-core-arts-standards.

Palmarini, J. (2011). The role of cornerstone tasks in the next generation of theatre standards: An interview with assessment expert Jay McTighe. Teaching Theatre. https://www.jaymctighe.com/wp-content/uploads/2011/04/Cornerstone-Tasks-in-Theater.pdf

Parker, K., & Igielnik, R. (2020). *On the cusp of adulthood and facing an uncertain future: What we know about Gen Z so far*. Pew Research Center: Social & Demographic Trends. www.pewresearch.org/social-trends/2020/05/14/on-the-cusp-of-adulthood-and-facing-an-uncertain-future-what-we-know-about-gen-z-so-far-2/

Saavedra, A. R., Liu, Y., Haderlein, S. K., Rapaport, A., Garland, M., Hoepfner, D., Morgan, K. L., Hu, A., & Lucas Education Research. (2021). *Project-based learning boosts student achievement in AP courses*. Lucas Education Research.

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., Bruner, J., Moeller, J., Linnansaari, J., Juuti, K., & Viljaranta, J. (2016). Investigating optimal learning moments in US and Finnish science classes. *Journal of Research in Science Teaching*, *53*(3), 400–421. https://doi.org/10.1002/tea.21306

Stefanou, C., Stolk, J. D., Prince, M., Chen, J. C., & Lord, S. M. (2013). Self-regulation and autonomy in problem- and project-based learning environments. *Active Learning in Higher Education*, *14*(2), 109–122. https://doi.org/0.1177/1469787413481132

Surahman, E., Wedi, A., Sulthoni, Soepriyanto, Y., & Setyosari, P. (2018). Design of peer collaborative authentic assessment model based on group project-based learning to train higher order thinking skills of students. *Advances in Social Science, Education, and Humanities Research*, *285*. International Conference on Education and Technology www.atlantis-press.com/proceedings/icet-18/125926623

Wiggins, G., & MiTighe, J. (2011). *The understanding by design to creating high-quality units*. ASCD.

CHAPTER

11

Accommodations and Modifications of Assessments for Twice Exceptional and English Language Learners

Yara N. Farah

Access is a fundamental aspect of instruction and assessment for all students. It is easy to recall a testing experience, whether for low or high stakes, where the test questions covered content that had not been taught. Many of us can remember assessment items that were unclear or poorly worded, thus making it almost impossible to show what we had learned about the content intended to be measured. Most of us would describe such assessments as unfair and non-representative of our learning. A large majority of us perhaps would develop negative attitudes towards assessment. In such situations, our performance would be negatively influenced. In fact, we did not have access to appropriate assessment. The purpose of assessment is to measure newly learned knowledge, skills, and concepts and verify attainment of learning objectives (Assessing Special Education Students [ASES] & State Collaborative on Assessment and Student Standards [SCASS], 2011). By improving access to instruction and assessment, the purpose of assessment can be met, and results can be meaningful, representative, and a positive experience for all students.

For many students with disabilities and those who are English language learners (EL), access to assessment has been constantly affected, if not denied (Thompson et al., 2018). Many experience limited opportunities to learn valued knowledge, skills, and concepts as well as to demonstrate learning. Valid claims about student performance cannot be made without appropriate access to instruction and assessment (Peltier & Harrison, 2018). Ideally, all students should be provided high-quality instruction and appropriate assessment. Assessment should be designed to provide all students with optimal access to targeted learning objectives without introducing variance due to extraneous test features (Lane et al., 2016). Access and the principle of equal protection is the basis of all federal legislation concerning education (Willner & Mokhtari, 2017). Thus, no educator involved in assessment today can afford not to know about the various needs of test takers, how they optimally participate in assessment tasks, and the implications of their participation.

DOI: 10.4324/9781003285991-16

The Movement towards Inclusion

The demographic makeup of the United States has changed dramatically over the past five decades. K-12 schools are serving students who are increasingly diverse in cultural background, language, socioeconomic status, and disability. Between 2011–12 and 2018–19, students receiving special education services under the Individuals with Disabilities Education Act (IDEA) increased from 13% of total public-school enrollment to 14% of total public-school enrollment (NCES, 2020). In the 2018–19 school year, 33% of these students had specific learning disabilities, 19% had speech or language impairments, and 15% had other health impairments (including those having limited strength, vitality, or alertness due to chronic or acute health problems). Moreover, the percentage of students served under IDEA was highest for American Indian/Alaska Native students (18%), followed by Black students (16%), White students and students of two or more races (14% each), and Hispanic students (13%).

English Language Learners (ELs) also constitute a notable percentage of public school students. In fall 2017, 5 million public school students (10.1%) were ELs (NCES, 2019) with the highest percentage in California (19.2%), followed by Texas (18%) and Nevada (17%). Spanish was the home language of 3.7 million EL public school students, representing 74.8% of all EL students and 7.6% of all public K–12 students. Arabic and Chinese were the next most commonly reported home languages (spoken by 136,500 and 106,500 students, respectively). English was the fourth most common home language for EL students (94,900 students), which may reflect students who live in multilingual households or students adopted from other countries who were raised speaking another language but currently living in households where English is spoken.

With these demographic changes over the past five decades, federal laws have also changed (Elliott et al., 2011). Policies have been developed to improve access to free and appropriate education for all students. Key federal legislation on access for students with disabilities includes the Elementary and Secondary Education Act of 1965 (ESEA), Section 504 of the Rehabilitation Act of 1973, and the Education for All Handicapped Children Act of 1975 (EHA) and their subsequent reauthorizations (Lane et al., 2016). These laws include students with disabilities in standards-based reform and test-based accountability under the No Child Left Behind Act (NCLB) of 2001 and require that assessments are selected and administered to ensure that results of testing accurately reflect the student's educational aptitude or achievement level.

However, these federal laws do not require states and districts to provide for the educational needs of gifted and talented learners (National Education Association, 2006). The Jacob Javits Gifted and Talented Students Education Act is the only federal program that was passed to support and address the needs of gifted and talented children. The Act focuses resources on identifying and serving students who are traditionally underrepresented in gifted and talented programs, particularly minority, economically disadvantaged, English language learners, and students with disabilities, to help reduce gaps in achievement and to encourage the establishment of equal educational opportunities for all students (U.S. Department of Education, 2019). However, the Act must be funded each year by Congress and does not fund local gifted education (NAGC, 2021). Decisions related to access for gifted learners, therefore, are made at the state and local levels.

Accommodations and Modifications in Assessments

Before any school personnel can implement any of the legally mandated test alterations, they must first understand the concept of *access,* and the difference between *accommodation* and *modification*. These three terms have been used for two decades when discussing educational assessments and the validity of resulting test scores (Elliott et al., 2011).

> For educational testing, *access* refers to opportunity for test-takers to demonstrate proficiency on the target construct of a test (e.g., language arts, mathematics, or science) or item (e.g., synonyms, homonyms, and homographs). In essence, complete access is manifest when a test taker is fully able to show the degree to which he or she knows the test content. Access, therefore, must be understood as an interaction between individual test-taker characteristics and features of the test itself. (Elliott et al., 2011, p. 3).

By examining the literature, it is clear there is a consensus that a*ccommodations* and *modifications* are access-enabling strategies. While these terminologies have been used interchangeably over the years, no clear definition has been provided in the *Standards for Educational and Psychological Testing*, and no explicit guidelines have been specified by federal or state laws (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Bruininks et al., 1994; Thurlow et al., 1993, 1996, Tindal & Haladyna, 2012). Both words—accommodation and modification—are used to express changes that are made to what is considered the "standardized" assessment conditions. To clarify the array of definitions, Hollenbeck (2002) specifies that test alteration moves along a continuum between accommodation and modification, depending on the degree of presence or absence of the following attributes:

- *Unchanged constructs*. Measurement of the construct should not be changed by the test alteration. When there is no alteration in the construct measured, then it is considered to be an accommodation.
- *Individual need*. Changes in the test should be completed to meet the student's needs. A change that is necessary and eliminates irrelevant factors is an accommodation as long as the content of the test is not altered.
- *Differential effects*. Changes in the test should affect a student's performance when compared with peer group performance. For the alteration to be considered an accommodation, it should not be useful for all individuals but rather focus on the targeted student.
- *Sameness of inference*. Alteration in tests should yield scores that could be compared with the standard administration scores. Accommodation does not alter the score of the test, and hence, the student's results can still be compared with the standard scores.

The more these four attributes are not in consensus, the more that the test alteration moves along the continuum from being an accommodation towards a modification (Hollenbeck, 2005). Their presence signifies an accommodation, whereas their absence signifies a modification. This continuum has been cited and adapted by many in the field of educational and psychological testing (Fincher, 2013; Kettler et al., 2018; Witmer et al., 2017).

Accommodations are mainly changes in *how* a student accesses information and demonstrates learning (Lane et al., 2016). While accommodations are intended to reduce or even eliminate the effects of a student's disability, they do not reduce learning expectations (Assessing Special Education Students & State Collaborative on Assessment and Student Standards, 2011). The changes are made in order to provide a student with equal opportunity to show what he or she knows and can do. The most frequent categories of testing accommodations and information on who can benefit include:

- *Presentation accommodations.* Students access information in ways that do not require them to visually read standard print. These alternate modes of access are auditory, multi-sensory, tactile, and visual.
- *Response accommodations*. Students complete activities, assignments, and assessments in different ways or solve or organize problems using some type of assistive device or organizer.
- *Setting accommodations*. The location in which a test or assignment is administered, or the conditions of the assessment setting, are changed.
- *Timing and scheduling accommodations*. The allowable length of time to complete an assessment or assignment is increased and the way the time is organized may be changed (Thurlow et al., 2005, p. 236).
- *Linguistic accommodations*. The cognitive resources ELs need to process the language of the test are minimized and the cognitive resources available for accessing the content of the test are maximized (Rivera & Collum, 2004, p. 3).

Table 11.1 provides information on who can benefit from accommodations, what questions to ask, and examples.

The type of accommodation any one child with a disability receives is based on an individual consideration of that child's needs. Consider the following two scenarios:

> *Scenario 1*
>
> Tenth grader Paul has been identified as a gifted student and English Language Learner (EL). His understanding and performance in literature exceeds his classmates when given appropriate accommodation such as presenting the information in multiple formats (i.e., written and orally). On the day of Paul's literature test, he arrives at school rested and confident of his ability to demonstrate his knowledge and skills on the test. Paul has already discussed his accommodation procedure with his teacher. He knows that the test reading passage and questions will be presented on a screen and will be read out loud using text-to-speech technology. Paul knows he will be able to control the speed, as well as raise or lower the volume of the voice. He knows that he can type his answers rather than handwrite his answers. The accommodation he will be using on the test is the same accommodation he usually uses in his classroom. When Paul arrives in the testing room, the test is on a large screen along with headphones, which he usually uses in class. He is greeted by his teacher who reviews one practice question to make sure all the equipment is working properly. While taking the test, Paul feels comfortable and confident of his ability to demonstrate his knowledge and skills and progresses from one question to the next smoothly. After he finishes his work, he leaves the room, and the teacher

TABLE 11.1 Accommodations for students with disabilities by category and area of need

Category of accommodation	Who can benefit	Questions to ask	Examples
Presentation	Students with print disabilities, defined as difficulty or inability to visually read standard print because of a physical, sensory, or cognitive disability.	Does the student have a visual impairment? Can the student read and understand directions? Does the student need directions repeated frequently? Has the student been identified as having a reading disability?	• Visual presentation: large print, reduce number of items per page or line, magnification devices, sign languages • Auditory presentation: audio tapes, screen reader/ text-to-speech, provide a designated reader • Tactile presentation: Braille, Nemeth code, tactile graphics • Talking materials: calculators; clocks, timers
Response	Students with physical, sensory, or learning disabilities (including difficulties with memory, sequencing, directionality, alignment, and organization)	Can the student use a pencil or other writing instrument? Does the student have a disability that affects his ability to spell? Does the student have trouble with tracking from one page to another and maintaining her place?	• Allow for answers to dictated to a scribe • Note-takers • Tape recorder • Respond on test booklet • Spelling and grammar devices • Graphic organizers • Allow for verbal responses • Permit responses to be given via computer
Setting	Students who need more time, cannot concentrate for extended periods, have health-related disabilities, fatigue easily, special diet and/or medication needs.	Can student work continuously during the entire time allocated for test administration? Does the student tire easily because of health impairments? Does student need shorter working periods and frequent breaks?	• Extended time • Frequent breaks • Multiple testing sessions
Timing and Scheduling	Students who are easily distracted in large group settings, concentrate best in small groups.	Do others easily distract the student? Does student have trouble staying on task? Does student exhibit behaviors that would disrupt other students?	• Provide preferential seating • Provide special lighting or acoustics • Provide a space with minimal distractions • Administer a test in small group setting • Administer a test in private room or alternative test site • Allow subtests to be taken in a different order • Administer a test at a specific time of day

(*continued*)

TABLE 11.1 Cont.

Category of accommodation	Who can benefit	Questions to ask	Examples
Linguistic	Students who are English language learners, non-native English speaker, or multi-lingual students.	Does the student direct more cognitive resources to process the language? Does the student interpret the test questions differently based on his/her culture? Does the student's understanding of the concept relate to the delivery language?	• Simplified English • Written in native language • Oral translation in native language • Use of bilingual dictionary • Allow responses in native language • Visual/picture presentation of material

prints his answers to be included in the pile of all tenth graders who are taking this test.

Scenario 2

Tayma is seven years old and loves school. She enjoys learning new concepts. Tayma has been identified as a twice-exceptional learner. Tayma can demonstrate her true knowledge when given questions one at a time with short rest breaks in between. Her performance level exceeds those of her peers even though she tires quickly. Today, Tayma is taking a math test. Her teacher already prepared the test to address her needs. Each question is given to her individually with five-minute breaks in between. Tayma feels very confident about her ability to answer the questions. She knows that taking the test in segments will provide her the opportunity to show her understanding of the math concepts. Tayma completes the test, knowing she did the best she could.

Both scenarios are examples of accommodations. For Paul and Tayma, the construct being assessed stayed the same. In fact, changes have been made so that the students can demonstrate their level of performance. Psychometrically, accommodation should remove construct-irrelevant variance in scores by providing access to demonstrate knowledge without interference of confounding influences such as test format, setting, or timing (Tindal & Haladyna, 2012). Accommodation offers the option of demonstrating academic knowledge without altering the validity or score of the test. To ensure test validity and score comparability, accommodations must not give a demonstrable advantage to students who receive them over students who do not (Elliott et al., 1998).

Test alterations that change the construct being assessed are considered modifications. Modifications are changes in *what* a student is expected to learn (Tindal & Haladyna, 2012). The changes are made to provide a student with opportunities to participate meaningfully and productively along with other students in classroom and school learning experiences. A modification includes a change in instructional level, content, curriculum, or performance criteria (Lane et al., 2016). For example, if the test calls for students to demonstrate two-digit multiplication, then altering the test so that the student demonstrates one-digit multiplication only would change the construct and would move the alteration toward a modification. Another modification example would be altering a silent-reading comprehension test (i.e., reading to oneself) to a listening comprehension test (i.e., test is read aloud). The standards being assessed in both examples have been changed requiring changes in evaluation and scoring. These types of modifications are typically done when a student cannot participate meaningfully in assessment without modification (Rivera & Collum, 2004). On the other hand, above-level modifications can also be made for gifted students. In the previous example a test that calls for students to demonstrate two-digit multiplication might be altered so that the student demonstrates three-digit multiplication. While this modification would change the construct and would move the alteration toward a modification, it would alter the content. In summary, accommodations do not change the expectations for learning and do not reduce the requirements of the assessment whereas modifications do change the expectations for learning and reduce or alter the assessment's requirements. Accommodations and modifications should meet the individual needs of each student.

Research Findings and Implications

Most of the extant research examining the effects of accommodations focus on the assessment performance of gifted students with disabilities (i.e., twice-exceptional students). These studies have addressed four discrete accommodations: time and scheduling, oral delivery, technology administration, and response accommodations.

Time and scheduling. Joakim (2015) and Ohleyer (2016) investigated 1.5 or 2 times the standard time provided for testing writing skills of K-12 students. The authors found that extended time did not affect the results on the writing test (Joakim, 2015; Ohleyer, 2016). Zentall et al. (2001) reported that twice-exceptional students did perform better when the assignment was short with detailed directions. Students were able to demonstrate their performance when the assessment was broken down into smaller tasks, and students were provided with checkpoints. Baum et al. (2014) also found that students with gifts and talents built stronger self-esteem and self-acceptance when extra time was provided, and they did not need to rush or put external demands on their time.

Oral delivery. Studies looking at oral delivery accommodation investigated the in-person read aloud, as well as voice recording and test reading software or text-to-speech devices. Kim (2016) found that comprehension with in-person versus recorded oral delivery had different impacts, depending on the grade level. For example, kindergarten and Grade 2 students scored better in comprehension with in-person versus recorded oral delivery, while Grade 4 students scored the same in comprehension in both oral delivery conditions. McMahon (2016) reported that all Grade 6 students scored significantly better in the oral delivery condition than without accommodation. Ohleyer (2016) found that students with learning disabilities in Grades 4, 5, and 6 performed better on writing assessments when using read-aloud directions only and when using assistive technology versus using no accommodation.

Technology administered. Seo and De Jong (2015) found no difference in performance for students in in Grades 6 and 9 when taking a social studies test presented in the traditional paper-based format compared to the test presented via computer. Eberhart (2015) investigated the effect of different technology on math and language art performance of Grade 7 students. The findings indicated that students scored statistically higher when the test was delivered via computer versus a tablet. Gunter and Kenny

(2012) found that the use of technology was particularly beneficial for twice-exceptional students. The use of technology made the assessment process more engaging, with dynamic real-time responses. The use of technology also aided in matching talents to the content, creating experiences of success (Wang & Neihart, 2015) and increased self-efficacy (Baum et al., 2014).

Response accommodations. Bouck and colleagues (2015) reported that middle school students with various disabilities performed better on math computation and word problems when using a graphing calculator accommodation. Higgins and colleagues (2016) compared math performance with and without American Sign Language (ASL) accommodations. Findings showed that students who were deaf scored on average consistently and significantly higher when using the accommodations at elementary, middle, and high school levels. Twice-exceptional students expressed a strong belief that using alternatives to written projects was an appropriate method to demonstrate true ability since many of them had difficulty with writing (Weinfeld et al., 2005). In addition, accommodations to written projects empowered and enabled students with twice exceptionalities to become more creative in expressing their thinking.

Some of the studies looking at accommodations focused on student perceptions. Researchers found that students had favorable impressions about specific accommodations—such as speech recognition tools (Nelson & Reynolds, 2015; Weis et al., 2016), and tactile graphics (Hansen et al., 2016)—or accommodations overall (Kafle, 2015; Ruhkamp, 2015; Timmerman & Mulvihill, 2015; Williams, 2015). Students preferred online testing (70%) over paper-based testing (10%), with 20% of students having no preference (Seo & De Jong, 2015). However, students were less likely to disclose information about their disabilities and seek accommodations when they had more negative views about seeking accommodations and more negative associations with their disabilities (Cole & Cawthon, 2015; Lyman et al., 2016; Ruhkamp, 2015). However, Weinfield et al. (2005) found that twice-exceptional students viewed accommodations positively as they allowed students to better demonstrate their knowledge as well as move from dependence to independence.

Other studies looked at teacher perceptions related to assessment accommodations. Educators reported that they felt that staff needed more access to assistive technology training (Ajuwon et al., 2016; Gallego & Busch, 2015), and that school districts prepared them to provide accommodations for students with disabilities more than their academic training programs (Detrick-Grove, 2016). Two studies indicated that educators had positive attitudes toward accommodations, and that teachers tended to report that low-tech accommodations—such as reading directions and reading test questions out loud—were more beneficial for students than more high-tech options (Detrick-Grove, 2016; DePountis et al., 2015). A study on perceptions related to accommodations for twice-exceptional students showed that general and special education teachers strongly believed that accommodations allowed students to participate more fully in assessments and to demonstrate knowledge (Weinfield et al., 2005).

Research to validate accommodations is growing, but this research is difficult to conduct, and the findings are often mixed. The effects from accommodations are highly influenced by each student, suggesting the importance of the individualized assignment of accommodations Both school personnel and policy makers struggle with decisions about which accommodations are needed and also maintain validity (Lane et al., 2016). Accommodations should aim to provide test-takers with appropriate conditions to demonstrate proficiency on the targeted concepts and skills (Elliott et al., 2011). In essence, the test needs to be representative of each test-taker's true level of performance and not irrelevant factors.

Practical Examples

The presence of Hollenbeck's (2002) attributes (i.e., unchanged constructs, individual need, differential effects, and sameness of inference) may be seen in the following scenarios. A discussion of each alteration selected by the teacher is discussed in respect to whether it is an accommodation or a modification, whether it fits the need of the individual student, and how the changes might affect the validity of the assessment.

Scenario A

Lana is a high school student who does not receive special education or Section 504 services but is accelerated in mathematics. The teacher has observed that Lana has strong critical thinking and problem-solving skills. When presented with problems to solve, she is able to clearly explain her thought process and justify her strategies; however, Lana often makes simple mistakes when multiplying or dividing large numbers with decimals. When she uses a calculator, Lana arrives at the correct answers. Her math teacher decided that Lana would be allowed to use the calculator during the problem-solving section of the exam, but not for the algebra section.

Student Data

- Accelerated in mathematics
- Ability to clearly explain her thought process
- Ability to justify her strategies
- Difficulty in hand multiplying or dividing large numbers with decimals
- Ability to use a calculator to multiply or divide large numbers with decimals

Alteration

Use a calculator during the problem-solving section of the exam but not for the algebra section.

Alteration Evaluation: Modification or Accommodation?

Since the problem-solving section of the exam assesses the student's ability to formulate hypotheses, select strategies, and communicate findings, the use of a calculator does not change the expectations for learning and is related to an appropriate accommodation for Lana. In fact, the calculator is a tool that will aid Lana in demonstrating her true performance on the assessment without changing the construct being measured. On the other hand, the algebra section of the exam assesses the student's ability to multiply and divide integers and correctly place decimal points in the solutions. Since student performance on this section helps determine if additional instruction is needed for calculating multiplication and division problems, the use of a calculator during the algebra section would alter the validity of the assessment—the learning expectations and the skills being measured. Hence, by solving this section of the exam without a calculator, Lana's true performance will be measured. In the long term, the teacher might decide to make further modifications by eliminating this objective for Lana and allow her to write out her thinking about the concept of place value. Using a calculator or writing out her thinking would be considered accommodations.

Scenario B

Carlos's second language is English. His teacher noticed that on his history test he is able to recount a series of events in detail and answer any questions after watching a documentary about the great depression. He shows his understanding of the chronology of events, assesses the impact of events on society, explains how events are influenced by historical development, and predicts long-term effects. However, Carlos finds it difficult to answer questions related to historical events that he has only read about. It is clear that Carlos's comprehension of historical events is affected by the mode of content delivery. Accordingly, the teacher decided that Carlos should watch a documentary or have access to visual graphics about the topic before any classroom test.

Student Data

- Second language is English
- Able to recount the series of events in details after watching a documentary
- Answers any question after watching a documentary
- Difficulty answering questions related to a historical event that he has read

Alteration

Watch a documentary or access visual graphics on the topic before any classroom test.

Alteration Evaluation: Modification or Accommodation?

The classroom history test requires students to read about a historical event and answer questions. The questions assess the student's ability to understand the chronology of events, explain how events are influenced by historical development, assess the impact of events on society, and predict long-term effects. By providing Carlos with a documentary on the topic, the teacher does not change the learning expectations but provides an opportunity to demonstrate his true performance in relation to comprehending historical events. In other words, the knowledge and skills being measured are unchanged and the alteration is based on Carlos's needs, which relate to his facility with English. Other possible accommodations might involve the use of visual graphics that reflect the chronology of events. Such alterations are considered accommodations.

Scenario C

Mirella is a first-grade student who has difficulty using a pen or pencil to write or draw. She also has strong geometric reasoning skills and is able to solve problems at a third grade level. During her free time, Mirella loves to measure the sides of the wooden tangram puzzles pieces and calculate the area of the shapes using a calculator. She knows how to measure the area of a triangle and a rectangle. She sometimes forms and calculates the area of a large rectangle, using smaller triangle and rectangle shapes. During recess, Mirella plays with Jasmine, a third-grade student. Jasmine draws shapes in the sand, and Mirella does the measuring. They discuss their thinking process and sometimes talk through the different possible strategies. In her geometry class, the learning activity assesses how students are learning to draw and distinguish the different attributes of polygons such as triangles and rectangles. Mirella is working with a note-taker to solve a project consisting of partitioning a large rectangle into four triangles with equal areas and two squares with equal areas.

Student Data

- Difficulty using a pen or pencil to write or draw
- Able to form larger shapes using small shapes
- Able to calculate the area of shapes by partitioning the shape
- Able to explain her thinking process orally
- Able to explain different strategies

Alteration

Work with a note-taker to solve a project consisting of partitioning a large rectangle into four triangles with equal areas and two squares with equal areas.

Alteration Evaluation: Modification or Accommodation?

The activity given to students assesses their ability to draw and distinguish different attributes of polygons. Mirella has already mastered this objective. Her geometric reasoning skills are two levels advanced at a third-grade level. The advanced activity given to Mirella assesses her ability to partition a large shape into specific smaller shapes, a more complex concept. The result of this activity will be differentiated from the one made for her classmates. This alteration is an above-level modification that provides Mirella with an opportunity to demonstrate her true performance in geometry. Because Mirella has difficulty drawing, she will work with a note-taker who will transcribe her reasoning that is shared orally. In this case, the alteration does not modify the construct being measured. It is based on Mirella's needs and will generate appropriate inferences about her level of performance. Using a note-taker when solving her geometry project is an appropriate accommodation.

Scenario D

Dave is a middle school gifted student with ADD. He is very creative in presenting his ideas during discussion and in his final products and performances; however, he has difficulty organizing his time and seldom completes assignments by the due date. His gifted education teacher has assigned a long-term project, researching the history of his school. The assignment requires that he formulate questions, interview three former graduates of the school, visit the school's library to review primary documents, and create a presentation about his research by the end of the first six weeks. The assessment rubric focuses on all aspects of the project. Because the teacher knows that Dave has difficulty completing long-term assignments, she breaks the assignment into smaller pieces (e.g., week 1—formulate questions; week 2—gather information from the school library; week 3—using the information, reformulate questions and create interview questions; week 4—interview three people who attended the school; week 5—organize information for a class presentation; week 6—present his information to the class). After each piece, she schedules a conference to review his progress on the assessment rubric.

Student Data

- Identified disability of ADD
- Creative in presenting ideas and presentations
- Difficulty in completing projects on time without support

Alteration

Break the assignment into smaller pieces and schedule a weekly conference to assess progress.

Alteration Evaluation: Modification or Accommodation?

The project requires the student to formulate questions, research the history of the school and present information to the class. This project is differentiated for a gifted class, given the emphasis on independent research and the process skills associated with that type of project. Yet it is the type of project that Dave has had difficulty with in the past because of procrastination in working productively across weeks. By breaking the project into smaller pieces and conferencing weekly with the student related to his progress, the teacher does not change the learning expectations but provides an opportunity for him to demonstrate his true performance and creativity in researching and presenting the project. These alterations are considered accommodations.

Challenges Related to Accommodations

Challenges related to accommodations can be categorized into three groups: policies, research, and school personnel decision making. The challenges for each of these groups are interrelated and influence one another as described in the following commentary.

Policies

Policies are important because they provide the framework for what occurs in school. The National Center on Education Outcomes has tracked the accommodation policies in all states and has provided information on their website (www.nceo.info). Their data show that all states recommend the use of assessment accommodations (Rogers et al., 2016). However, since accommodations are not clearly defined, limited consensus exists on what constitutes appropriate accommodations and almost no information on how to interpret and report the scores for students who complete an assessment with accommodations. Since the 1990s, however, changes have occurred in state policies. Policies are now focusing more on the need to clarify the purpose of the test, the construct being tested, and the importance of providing appropriate access to assessment (Lane et al., 2016). Unfortunately, this progress is different between states. What is allowed in one state might be prohibited in another state. This is problematic for students moving from one state to another. The issue of clear and consistent policies begs for a solution that is evidence-based and is connected to the next challenge related to accommodations.

Research

Although accommodation policies, especially in the K–12 system, have been studied for many years, it is only relatively recently that research has existed on the use of accommodations, particularly with gifted students (Thurlow, 2001). Research indicates that there is considerable variability from state to state in the percentages of students with disabilities using accommodations, which range from 8 to 82% (Thompson & Thurlow, 1999). There is also a disconnect between the accommodations most frequently permitted in state policies and those that are most frequently used. Because of these inconsistencies in definition, results from studies do not provide clear implications. In fact, researchers, those designing tests, and practitioners grapple with decisions about the role of accommodations in providing access to assessment opportunities (Lane et al., 2016). This brings the field into a cycle of policies that need to be justified with evidence-based research, and research needed to study the validity of what is proposed in policies. For this reason, there is lack of consistency in both policies and research. This situation highly affects the decision-making process of school personnel.

Decision Making of School Personnel

The process of making decisions about accommodations and modifications and then ensuring those decisions are carried out is a major challenge. School personnel are responsible for determining appropriate accommodations and above-level modifications of assessments frequently during IEP and 504 meetings (Lane et al., 2016). However, no clear guidance is provided by policies nor through research, which makes it even more important for educators to be aware of the similarities and differences so they might advocate for twice-exceptional and gifted students who are English learners. Although these implementation difficulties can be improved through professional learning, there continues to be a lack of training in pre-service education and state training on accommodations and modifications (Rogers et al., 2016). The development of a decision-making strategy and adequate training of school personnel, as well as solutions to many logistical problems that surround the provision of accommodations and modifications in K–12 assessments, are essential for moving toward better assessments (Elliott et al., 2011).

Recommendations for Coordinators and Directors of Gifted Education Programs

Coordinators and directors of gifted education programs should organize, promote, and support school personnel in using assessments for identification and learning progress. The National Association for Gifted Children (NAGC) developed Pre-K-to Grade 12 Gifted Education Programing Standards with the aim to assist school districts in examining the quality of their programs and services for gifted learners. Programing Standard 2 focuses on assessment and states:

> Knowledge about different uses of assessment is essential for educators of students with gifts and talents. It is important to understand assessments when assessing abilities and achievement, designing services, and identifying students in need of services, and assessing each student's learning progress. … Educators' understanding of technically adequate and equitable approaches that minimize bias will enable them to select and use the assessment tools needed to identify students who represent diverse backgrounds. … As a result of each educator's use of ongoing assessments, students with gifts and talents are aware of their learning progress and demonstrate growth commensurate with their abilities. (National Association for Gifted Children, 2019, p. 9)

This standard offers some powerful concepts that parallel federal laws, research findings, and the literature related to assessment accommodation.

The standard articulates that knowledge about technical adequacy and different uses of assessment are crucial for educators of students with gifted and talents. The standard expresses the importance of "using equitable approaches that minimizes bias" and that "demonstrate growth commensurate with their abilities." This emphasis aligns with the concept of *access.* As mentioned earlier in this chapter, "access refers to opportunities for test-takers to demonstrate proficiency." Educators must face the challenge of selecting appropriate accommodations that preserve test validity, are comparable across test scores, and provide opportunity for gifted students with disabilities to demonstrate proficiency. Currently, several forms of accommodations and procedures are suggested in this chapter and can be used as a beginning point.

The ultimate goal is to provide all students *access to appropriate assessments*. Because policies, research base, and training in the field is limited, the school district will need to (a) define accommodations and above-level modifications; (b) review, summarize, and share current research findings on assessments with gifted learners who have disabilities or are English language learners; (c) identify and provide resources for assessment accommodations and above-level modifications; (d) evaluate the district's implementation of assessment accommodations and above-level modifications with gifted students who are English learners or who have disabilities; (e) develop a plan of action for improving the use of assessments with gifted learners who have special needs; and (f) promote assessment as a way to effectively address the needs of students.

Final Thoughts

The important ideas concerning assessment accommodations and above-level modifications have been identified in this chapter. There are several crucial points to keep in mind. First, more research is needed; however, some of the critical implementation questions are not likely to be answered by randomized experimental research designs since assessment accommodations and above-level modifications are supposed to be effective only when they meet an individual student's needs. A better approach might be to build consensus among experts about which test accommodations and above-level modifications are necessary to support effective practices. Second, it is essential for educators to learn about national standards related to assessments and the principles behind accommodations and above-level modifications. Third, students participating in assessments should understand the role of accommodations and above-level modifications. Once students move beyond elementary school, some students become concerned about "being different" from other students and refuse to use accommodations that are probably needed. Students need to fully understand the consequences of this choice. Finally, the essential role that access plays in achieving valid assessment results can only be realized by making sound decisions about how to design and administer tests for students.

References

Ajuwon, P. M., Meeksa, M. K., & Okunga, P. A. (2016). Reflections of teachers of visually impaired students on their assistive technology competencies. *Journal of Visual Impairment & Blindness*, *110*(2), 128-134. https://doi.org/10.1177/-145482X1611000207

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Assessing Special Education Students & State Collaborative on Assessment and Student Standards. (2011). *Accommodations manual: How to select, administer, and evaluate. Use of accommodations for instruction and assessment of students with disabilities* (3rd ed.). Council of Chief State School Officers.

Baum, S. M., Schader, R. M., & Hébert, T. P. (2014). Through a different lens: Reflecting on a strengths-based, talent-focused approach for twice-exceptional learners. *Gifted Child Quarterly*, *58*(4), 311–327. https://doi.org/10.1177/0016986214547632.

Bouck, E. C., Bouck, M. K., & Hunley, M. (2015). The calculator effect: Understanding the impact of calculators as accommodations for secondary students with disabilities. *Journal of Special Education Technology*, *30*(2), 77-88. https://doi.org/10.1177/0162643415617371

Bruininks, R., Gilman, C., McGrew, K., Scott, D., Shriner, J., Spande, G., … Ysseldyke, J. (1994). *Testing accommodations for students with disabilities: A review of the literature (Synthesis Report 4)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Cole, E. V., & Cawthon, S. W. (2015). Self-disclosure decisions of university students with learning disabilities. *Journal of Postsecondary Education and Disability*, *28*(2), 163–179. https://files.eric.ed.gov/fulltext/EJ1074663.pdf

DePountis, V. M., Pogrund, R. L., Griffin-Shirley, N., & Lan, W. Y. (2015). Technologies that facilitate the study of advanced mathematics by students who are blind: Teachers' perspectives. *International Journal of Special Education*, *30*(2), 131–144. https://eric.ed.gov/?id=EJ1094788

Detrick-Grove, T. L. (2016). Teachers' perceptions and knowledge of accommodations for students with disabilities. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 78/05(E).

Eberhart, T. (2015). A comparison of multiple-choice and technology-enhanced item types administered on computer versus iPad. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 77/06(E).

Elliott, S. N., Kettler, R. J., Beddow, P. A., & Kurz, A. (2011). *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy*. Springer Science & Business Media.

Elliott, S. N., Kratochwill, T. R., & Schulte, A. G. (1998). The assessment accommodations checklist: Who, what, where, when, why, and how? *TEACHING Exceptional Children*, *31*(2), 10–14. https://doi.org/10.1177/004005999803100202

Fincher, M. (2013). *Investigating the effects of a read-aloud alteration on the third-grade reading Criterion-Referenced Competency Test (CRCT) for students with disabilities* [Doctoral dissertation, Georgia State University]. ProQuest Dissertations Publishing.

Gallego, M., & Busch, C. (2015). Towards the inclusion of students with disabilities: Accessibility in language courses. *Innovative Higher Education*, *40*(5), 387–398. https://doi.org/10.1007/s10755- 015-9321-z

Gunter, G. A., & Kenny, R. F. (2012). UB the director: Utilizing digital book trailers to engage gifted and twice-exceptional students in reading. *Gifted Education International*, *28*(2), 147–160. https://doi.org/10.1177/0261429412440378

Hansen, E. G., Liu, L., Rogat, A., & Hakkinen, M. T. (2016). Designing innovative science assessments that are accessible for students who are blind. *Journal of Blindness Innovation and Research*, *6*(1). https://doi.org/10.5241/6-91

Higgins, J. A., Famularo, L., Cawthon, S. W., Kurz, C. A., Reis, J. E., & Moers, L. M. (2016). Development of American Sign Language guidelines for K-12 academic assessments. *Journal of Deaf Studies and Deaf Education*, *21*(4), 383–393. https://doi.org/10.1093/deafed/enw051

Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessment. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students* (pp. 395–425). Lawrence Erlbaum.

Hollenbeck, K. (2005). Validity issues and decisions about testing accommodations. *Assessment for Effective Intervention*, *31*(1), 7–17. https://doi.org/10.1177/073724770503100102

Joakim, S. E. (2015). Help me fail: A study on testing accommodations for students with disabilities in writing assessments. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 77/03(E).

Kafle, E. A. (2015). Nursing students with learning disabilities: Perceptions and attitudes regarding the role of disability support program services and access to accommodations. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 76/08(E).

Kettler, R. J., Elliott, S. N., Beddow, P. A., & Kurz, A. (2018) Accessible instruction and testing today. In S. Elliott, R. Kettler, P. Beddow, & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices* (pp. 1–16). https://doi.org/10.1007/978-3-319-71126-3_1

Kim, Y. S. G. (2016). Do live versus audio-recorded narrative stimuli influence young children's narrative comprehension and retell quality? *Language, Speech, and Hearing Services in Schools*, *47*(1), 77–86. https://doi.org/10.1044/2015_LSHSS-15-0027

Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of Test Development* (2nd ed.). Routledge.

Lyman, M., Beecher, M. E., Griner, D., Brooks, M., Call, J., & Jackson, A. (2016). What keeps students with disabilities from using accommodations in postsecondary education? A qualitative review. *Journal of Postsecondary Education and Disability*, *29*(2), 123–140. https://eric.ed.gov/?id=EJ1112978

McMahon, D., Wright, R., Cihak, D. F., Moore, T. C., & Lamb, R. (2016). Podcasts on mobile devices as a read-aloud testing accommodation in middle school science assessment. *Journal of Science Education and Technology*, *25*(2), 263–273. https://doi.org/10.1007/s10956-015-9591-3

National Association for Gifted Education. (2019). *Pre-K to Grade 12 Gifted Programming Standards.* www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20Standards%281%29.pdf

National Association for Gifted Education. (2021). *Jacob Javits Gifted & Talented Students Education Act.* National Association of Gifted Children.

National Center for Education Statistics at Institute of Education Sciences. (2020). *The condition of education 2020* (NCES 2020-144). U.S. Department of Education.

National Education Association. (2006). *The twice- exceptional dilemma.* National Education Association Professional Library.

Nelson, L. M., & Reynolds, T. W. (2015). Speech recognition, disability, and college composition. *Journal of Postsecondary Education and Disability*, *28*(2), 181–197. https://eric.ed.gov/?id=EJ1074665

Ohleyer, A. A. (2016). Elementary tech: Assistive technology, specific learning disability, and state standardized testing. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 77/08(E).

Peltier, C., & Harrison, J. R. (2018). Selecting accommodations for mathematics assessments: Legal and practical considerations. *Preventing School Failure: Alternative Education for Children and Youth*, *62*(4), 300–310. https://doi.org/10.1080/1045988X.2018.1443425

Rivera, C., & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English language learners.* The George Washington University Center for Equity and Excellence in Education.

Rogers, C. M., Lazarus, S. S., & Thurlow, M. L. (2016). *A summary of the research on the effects of test accommodations: 2013-2014 (NCEO Report 402).* National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Report402/default.htm

Ruhkamp, R. (2015). Lived experiences of undergraduate and graduate students utilizing accommodations. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 76/11(E).

Seo, D. G., & De Jong, G. (2015). Comparability of online- and paper-based tests in a statewide assessment program: Using propensity score matching. *Journal of Educational Computing Research*, *52*(1), 88–113. https://doi.org/10.1177/0735633114568856

Tindal, G., & Haladyna, T. M. (2012). *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation.* Routledge.

Thompson, W., Coleman, J. M., Riley, K., Snider, L. A., Howard, L. J., Sansone, S. M., & Hessl, D. (2018). Standardized assessment accommodations for individuals with intellectual disabilities. *Contemporary School Psychology*, *22*, 443–457. https://doi.org/10.1007/s40688-018-0171-4

Thompson, S. J., & Thurlow, M. L. (1999). 1990 state special education outcomes: A report on state activities at the end of the century. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L. (2001). Use of accommodations in state assessments: What databases tell us about differential levels of use and how to document the use of accommodations [Technical Report 30]. Minneapolis: University of Minnesota, National Center on Educational Outcomes. https://eric.ed.gov/?id=ED457604

Thurlow, M., Hurley, C., Spicuzza, R., & Erickson, R. (1996). *Resources: Students with disabilities in national and statewide assessments (Minnesota Report No. 7).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Lazarus, S. S., Thompson, S. J., & Morse, A. B. (2005). State policies on assessment participation and accommodations for students with disabilities. *The Journal of Special Education*, *38*(4), 232–240. https://doi.org/10.1177/00224669050380040401

Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature (Synthesis Report 4).* Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Timmerman, L. C., & Mulvihill, T. M. (2015). Accommodations in the college setting: The perspectives of students living with disabilities. *The Qualitative Report*, *20*(10), 1609–1625. https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=2334&context=tqr

U.S. Department of Education. (2019). *Jacob Javits Gifted & Talented Students Education Act.* Office of Elementary and Secondary Education. CFDA Number 84.206A www2.ed.gov/programs/javits/index.html

Wang, C. W., & Neihart, M. (2015). Academic self-concept and academic self-efficacy: Self-beliefs enable academic achievement of twice-exceptional students. *Roeper Review*, *37*(2), 63–73. https://doi.org/10.1080/02783193.2015.1008660

Weinfeld, R., Barnes-Robinson, L., Jeweler, S., & Shevits, B. R. (2005). Enabling or empowering? Adaptations and accommodations for twice-exceptional students. *Teaching Exceptional Children Plus*, *2*(1), Article 6. https://files.eric.ed.gov/fulltext/EJ966532.pdf

Weis, R., Dean, E. L., & Osborne, K. J. (2016). Accommodation decision making for postsecondary students with learning disabilities: Individually tailored or one size fits all? *Journal of Learning Disabilities*, *49*(5), 484–498. https://doi.org/10.1177/0022219414559648

Williams, A. D. (2015). Middle school students' experience of receiving test accommodations. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 76/12(E).

Willner, L. S., & Mokhtari, K. (2017). Improving meaningful use of accommodations by multilingual learners. *The Reading Teacher*, *71*(4), 431–439. https://doi.org/10.1002/trtr.1637

Witmer, S., Schmitt, H., Clinton, M., & Mathes, N. (2017). Accommodation use during content area instruction for students with reding difficulties: Teacher and student perspectives. *Reading & Writing Quarterly*, *34*(2), 174–186. https://doi.org/10.1080/10573569.2017.1382407

Zentall, S. S., Moon, S. M., Hall, A. M., & Grskovic, J. A. (2001). Learning and motivational characteristics of boys with AD/HD and/or giftedness. *Exceptional Children*, *67*(4), 499–519.

CHAPTER

12 Student Self-Assessment

Anne N. Rinn and April Walker

Student Self-Assessment

Defining Student Self-Assessment

Student self-assessment "generally involves a wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products" (Panadero et al., 2016, p. 804). The study of student self-assessment has been a main area of research in the fields of education and educational psychology since the late 1980s (Boud & Falchikov, 1989; Falchikov & Boud, 1989). Researchers and practitioners seem to agree: The use of student self-assessment in the classroom is a fundamental component to learning and performance, and to the development of self-regulation and self-efficacy. Teachers can deliberately cultivate the development of self-regulation and self-efficacy through the use of student self-assessment, both of which will further contribute to learning and performance and the development of talent over time (see psychosocial skills; Subotnik et al., 2011, 2018) and are part of the larger notion of metacognition (Flavell, 1979).

Types and Characteristics of Student Self-Assessment

In a review of the literature on self-assessment, Panadero et al. (2016) suggest that the use of student self-assessment can happen in different ways and with various lenses (see Boud & Brew, 1995; Taras, 2010). Early models of self-assessment in the research literature distinguished between self-assessment and other methods of self-evaluation such as self-testing or self-marking (checking one's performance against provided items) and reflective questioning (prompting to help students better understand what they are reading) (see Boud & Brew, 1995). Later models incorporate multiple methods of self-evaluations in defining self-assessment.

For example, Panadero, Alonso-Tapia, and colleagues developed a typology of student self-assessment that includes three student self-assessment formats based on the presence and format of the assessment criteria (Alonso-Tapia & Panadero, 2010; Panadero et al., 2013). The three formats include:

(a) *standard self-assessment* (sometimes called self-grading) in which students are asked to self-assess without being given explicit criteria (most of the empirical research

DOI: 10.4324/9781003285991-17

using standard SSA [standard self-assessment] does not state whether, and if so how, criteria were provided);

(b) …*rubrics* in which students are given a rubric that includes criteria and performance standards with specific examples of the final product; and

(c) …*scripts* which include criteria presented as questions that the students need to answer for themselves; these are similar to prompts but focus on the task process. (Panadero et al., 2016, p. 808)

As another example, Brown and Harris (2013) classified student self-assessment according to how the self-assessment is carried out:

(a) *Self-rating* involves a rating system (e.g., a checklist of tasks or processes completed) that is used by students to judge the quality or some quantity aspect of their work.

(b) *Self-marking* involves having students grade their own work using objective scoring guides (e.g., a list of correct answers). These guides allow students to grade their work against an agreed upon standard or criterion.

(c) *Criteria- or rubric-based assessments* involve having students judge or evaluate their own work against descriptions of increasing quality. With this type of assessment, the focus is on using the rubric to guide judgment of quality and not just accuracy.

(d) *journaling* involves student reflection on what was created, why, and what else the student notices about the process of learning. A rubric or other tool may be used to judge the response, or more holistic assessment processes may be applied.

There is overlap in the models of student self-assessment seen here, but the models of student self-assessment all include components of self-marking and evaluation of a student's product and processes used to create it, and they can all be used as components of formative and summative assessment.

Multiple forms of self-assessment can be beneficial for students (Brown & Harris, 2013; Panadero et al., 2012). Self-assessment should be used in conjunction with pre-assessment, formative assessment, and summative assessment in all of the content areas. For example, using curriculum compacting, students can be given enrichment opportunities such as a project. Students given enrichment opportunities can be coached to use self-assessment throughout the learning process. A checklist of items or a learning contract (a *self-rating assessment*), can be provided to students to complete throughout a unit to have them report on their progress. In addition, students can be asked to respond to a *script* that models the type of desired reflection.

Purpose of Student Self-Assessment

Student self-assessment can be used as a method of formative assessment, for example, or on-going assessment, so that teachers gather and interpret evidence about a student's performance and then make decisions regarding differentiating instruction in order to improve student learning and achievement (Sadler, 1989). While teachers may use student self-assessment to improve teaching and instruction, students can also use self-assessment not only to improve their performance but also learn about their self-beliefs and their own thinking such as and how they go about solving problems (e.g., planning, monitoring, and evaluating). All of these involve metacognition, or a student's ability to think about their own thinking (Flavell, 1979). Metacognition is

> widely viewed as a critical hallmark of expert performance in that experts organize greater amounts of knowledge in a more effective manner, use more appropriate strategies, and regulate their thinking and performance more effectively than nonexperts. Some researchers in the field of gifted education have hypothesized that gifted individuals are distinct from their typical peers because they think like experts, particularly within the individual's area of exceptional ability. (Snyder et al., 2011, p. 181)

Student self-assessment should thus be viewed as an important component of metacognition and the learning and assessment process (Brown & Harris, 2013; Panadero et al., 2017). Students construct meaning, or learn, in part, by self-assessing prior to and during learning. Students connect new information, understandings, and skills with information they have already stored and used. Students then monitor, make connections, evaluate and internalize learning independently, enhancing learning in a meaningful, rather than rote, manner. We describe the impact of student self-assessment on learning and academic performance and on the development of self-beliefs as follows.

Impact on Learning and Academic Performance

Students who are trained to use student self-assessment typically experience an increase in learning and academic performance (Panadero et al., 2012; Topping, 2003). In a review of 23 studies on self-assessment that included a variety of types of self-assessment, Brown and Harris (2013) found student self-assessment positively impacted learning and academic performance across a range of grade levels and subject areas. They found that the type of self-assessment was less relevant than the use and complexity of a self-assessment. For example, in a study comparing the use of scripts in one group, rubrics in one group, and no self-assessment in the control group, Panadero et al. (2012) found the use of both scripts and rubrics enhanced learning compared to the control group.

Feedback further impacts the relationship between self-assessment and learning. In a review of research, Sitzmann et al. (2010) found the relationship between self-assessment and learning was stronger for courses that included feedback from the teacher than for courses that did not include feedback. In order for feedback to be most effective, it should be clear, it should build trust between the teacher and the student, it should be respectful, it should be specific, it should be differentiated, it should be timely, and it should invite follow-up (Tomlinson & Moon, 2013). For example, teachers can use live or recorded verbal feedback as an efficient way to provide timely, specific, and differentiated feedback that also builds relationships with students. It can be followed up with written comments on assignments for further feedback.

Impact on Self-Beliefs

Students trained to use self-assessment are likely to experience an increase in self-beliefs, in particular self-regulated learning strategies and self-efficacy. These are important, particularly as they further impact learning and performance, and are also important as psychosocial skills that contribute to the development of talent (Subotnik et al., 2011, 2018).

Self-Regulation. Self-regulation involves comparing one's current state with a target state (or goal) and then exhibiting motivated behaviors to move closer or entirely to the target state (Zimmerman & Schunk, 2011). Further, when tasks are tied to students' interests and real-world applications, there is likely to be greater motivation and evidence of self-regulation (VanTassel-Baska & Brown, 2007). For example, if students are asked to complete a performance-based task to build a model in a math class, they may be more motivated to understand the geometry and calculation skills needed to design and build the model (especially if it is a model of something in which they are interested). An example in a language arts class would be to ask students to write a persuasive letter to an administrator requesting a change in policy that would benefit students. The student may be more motivated to learn the persuasion techniques needed to write a convincing letter if there is the potential to see real-world benefits. Multiple models and theories of self-regulated learning exist (see Panadero, 2017). Self-regulation has been discussed as an important psychosocial skill in the development of academic talent (among other domains of talent; see Oppong et al., 2019; Subotnik et al., 2011, 2018).

Training in student self-assessment can help students regulate their own learning, and the relationship between self-assessment and self-regulation is fairly established in the research literature. In a meta-analytic review of 19 research studies, for example, Panadero et al. (2017) found consistent evidence for the effects of student self-assessment on self-regulation. Further, involving students in planning for self-assessment can further impact learning strategies. For example, involving students in the co-creation of a self-assessment rubric can lead to higher levels of learning self-regulation (Fraile et al., 2017).

Self-efficacy. In Bandura's (1986) social cognitive theory, individuals are viewed as proactive and self-regulating. How people behave, perform, and achieve can be predicted by the beliefs individuals hold about their capabilities. Bandura called these self-beliefs *self-efficacy*. Individuals tend to engage in tasks in which they feel competent and confident, or when they have feelings of self-efficacy regarding a task. They avoid those in which they do not. Self-efficacy and self-regulation are clearly related, as one must feel confident enough to try and move toward a target state (or goal).

Academic self-efficacy and academic achievement are almost always related in research studies (Burns et al., 2020; Hwang et al., 2016), and this relationship also occurs among samples of individuals identified as gifted and talented (Dixson et al., 2016). Self-efficacy matters at all ages and stages of development and has an impact on learning, performance, and talent development. Self-efficacy plays an important role in developing an interest in a career field, as well as eventually choosing a career and performing in that career (see Lent & Brown, 2017).

Using student self-assessment can help students develop feelings of self-efficacy related to the task at hand. In the meta-analytic review of 19 research studies mentioned earlier, Panadero et al. (2017) found strong evidence for the effect of student self-assessment on self-efficacy. Andrade et al. (2009) found similar results with an increase in self-efficacy ratings over time, particularly for girls who used rubrics.

Factors Influencing the Accuracy of Self-Assessment

Some research on student self-assessment examines factors that impact self-assessment accuracy such as ability level, age/grade level, and various environmental factors. For example, student ability level or achievement level has been shown to affect self-assessment accuracy. Research has shown that, even when using a scoring rubric, undergraduates receiving higher grades from professors are more accurate in self-assessment than students receiving lower grades (González-Betancor et al., 2019). In another study of college freshmen, students judged as being of higher academic ability were able to self-assess with higher accuracy than their peers with lower academic ability (Lew et al., 2010). This phenomenon, known as the Dunning-Kruger effect, occurs more often in lower-performing students (Kruger & Dunning, 1999). These students will not only perform poorly, but they will fail to recognize their poor performance and consequently do little to positively impact future learning (Dunning et al., 2003). This lack of metacognitive ability prevents students from not only succeeding at an academic level, but also hinders the self-regulated learning that is needed to be a lifelong learner, capable of adapting to any learning situation. As these studies were conducted on undergraduates, more research is needed to determine how ability levels impact self-assessment accuracy during the elementary and secondary years.

Further, age and overconfidence, or overestimating one's ability, may be inversely related such that older students are more accurate in self-assessment than younger students (Machado & Yoshinaga, 2018). With greater maturity, training, and practice, students are able to evaluate themselves more accurately as they age. This points again to the need for training in student self-assessment, particularly in the earlier grades.

Practical Examples of How Self-Assessment Might Be Used in School Settings

Coaching students to use self-assessment can be an integral component to building a differentiated curriculum. VanTassel-Baska and Little (2017) discuss that an important component of a differentiated curriculum is varying the curriculum and learning opportunities for gifted learners through the use of advanced content, higher level processes, and products, and the use of macro concepts. Further, deferential differentiation involves considering students' interests and preferred methods of learning (Kanevsky, 2011). Thus, when designing curriculum for advanced students, units can be built with choice and learning preferences in mind. Teachers can use interest inventories to get to know students and gain insight into student preferences. Students' interests could be incorporated by differentiating the content for students and/or by giving product choices. Learning contracts, such as those used in the units designed by Tomlinson and Eidson (2003), can be another way to use deferential differentiation to incorporate self-assessment. Students can have choice in the process described in their learning contract and use self-assessment to monitor their progress.

Students can also be asked to set goals throughout a unit and to help plan the pacing. They can even help create the rubrics for their own projects (and self-assess according to the criteria they developed). To illustrate, the second author of this chapter was teaching a unit on mock trial and argumentation with a group of advanced middle school students. The students were asked to help create the roles and rubrics for the unit. Students were put into groups and defined the requirements for each role (e.g., lawyer, witnesses, defendant), including the requirements to prepare for the trial and the performance during the trial. Students then created rubrics in their groups for how the performance of each role would be assessed during the trial and how the assignments created in preparation would be assessed.

Self-Assessment with Portfolios

Portfolios are another way for students to self-assess what products show evidence of their best work and thinking. The portfolio could be in one particular content area or across content areas. Siegle (2002) has discussed the potential of having students create electronic portfolios. Students can more easily store and organize examples of their work. Recordings of performances and presentations can also be stored for students to save throughout their educational career. Portfolios could also be kept in specific content areas. For example, students in a language arts class can keep a portfolio of their writing throughout the year and choose their best pieces for inclusion. Encouraging students to take an active role in selecting the best examples of their work to keep for the future helps them feel positive about their academic growth. It also demonstrates their understanding of both the quality of their own work and how that work may improve if they are required to provide a written reflection of their reactions to it.

Self-Assessment in Language Arts

Language arts is a content area that is rich with opportunity for self-assessment. For example, if students are reading novels in small groups such as literature circles, rather than the teacher providing the pacing for the reading and assignments, students can be provided a calendar and asked to set their own goals. They can have weekly group check-ins to see if they are meeting their goals (or need to adjust their goals). Giving students autonomy over their own learning and offering choice of pacing can be very motivating.

As another example, rubrics are very commonly used in language arts for assessing projects and writing. Students can be coached to *self-assess* with a rubric, measuring their own performance against a specific criterion. Figure 12.1 illustrates a self-rating type of checklist and reflection (Brown & Harris, 2013) that students could complete on their own writing before the teacher grades it with a rubric. While this checklist is intended for use in Grade 7, it could be adapted for other grade levels as well. The checklist was created, based on the seventh grade Common Core standards (CCSS) for Informative and Explanatory writing (Common Core State Standards Initiative, 2021), and represents a tool that is successfully applied for multiple purposes in programs for the gifted. After the second author observed students who were making many of the same mistakes with each essay, she created this rubric. Training students to complete the checklist before meeting with the teacher for writing conferences encouraged them to become more thoughtful about their own editing and revisions. Over time, students internalized the revisions they made to their writing and became stronger writers with fewer editing and revisions needing to be made.

Students can also use this checklist to *self-reflect* if they are completing the editing/revising process of an essay. This type of tool could also be used by a teacher during writing conferences to offer *tailored feedback* to students after the students self-assess. To help students be self-reflective, teachers should model the thinking process required when completing this checklist with a piece of writing. Using modeling to aid in self-assessment is supported by Bandura's (1986) social cognitive theory. Modeling the thinking process and using an example mentor text can help students build self-efficacy when they understand how to more accurately evaluate their own work.

Collaborative Analysis of Errors. Another way to scaffold the self-assessment process and help students learn to improve in their skills for revising is to provide students with

Name: ____________________________________ Date: _________________

Paper Title: ____

Checklist for Informative/Explanatory Writing

Directions: After completing your rough draft of your paper, read through this checklist. Ask yourself how you addressed each checklist item in your paper. Answer the questions here in the "Your Response" column. Go back and make revisions to your paper as needed.

Checklist Questions	Your Response
Introduction: What kind of hook did you use to grab the reader's attention?	
Introduction: Did you introduce the topic of the paper? (i.e., Can the reader tell what the body paragraphs will be about from the introduction?)	
Introduction: Did you include a thesis statement, and does it preview the topics of your body paragraphs?	
Organization: What organization strategy did you use to organize the topic of your paper (e.g., definitions, classification, comparison/contrast, and cause/effect)?	
Transitions/Flow: Did you include transition words to help your paper flow and show how the different topics discussed are related?	
Transitions/Flow: Is your paper cohesive? (i.e., Do your paragraphs make sense, stay on topic, and flow well between topics?)	
Topic Development: Did you use several supporting details in each of your body paragraphs to develop your points?	
Topic Development: When providing evidence for your points, did you use quotations or paraphrasing and give proper credit to the source? Did you introduce your quotes?	
Topic Development: Did you check that the quotes you chose are the best quotes to make the points you are trying to make? (i.e. Do the quotes make sense in the paragraph?)	
Topic Development: Did you include analysis to explain the meaning or significance of quotes?	
Conventions: Did you use proper formatting (i.e., MLA) for citing your quotes?	
Language/Style: Did you use appropriate vocabulary (i.e., technical terms) to explain the informative topics you discussed?	
Language/Style: Did you use a formal tone throughout the paper? (e.g., avoid contractions, using words like "you," and using slang).	
Conventions: Did you check spelling and grammar?	
Conventions: Did you check capitalization and punctuation?	
Conventions: Did you check that you do not have any run-on sentences or sentence fragments?	
Conventions: Did you check that margins, font, spacing follow formatting guidelines?	

Rough Draft Reflection

1. What went well in this paper?
2. What do you think you could have done better?
3. What was the hardest part about the research paper process?
4. What was the most helpful part of the research process?
5. What do you think could have made the process easier to understand and carry out?
6. What revisions do you need to work on in the final draft?

FIGURE 12.1 Sample checklist and reflection for a writing assignment

a teacher-created sample paper that needs edits/revisions. Students would work with a peer to go through the checklist and make revisions to the sample paper, using the items in the checklist. By working with a peer, they can collaborate and discuss the improvements. Students could then use the checklist on their own writing draft after practicing with the sample paper. The class could also have a whole-group discussion about what improvements each group made to the sample paper. The self-assessment process for writing could be taken a step further by having students self-assess their essay with a rubric after completing the checklist and revising their draft. See Figure 12.2 for an example of a rubric tied to the seventh grade CCSS (Common Core State Standards Initiative, 2021) Students can go through their paper and score themselves on each category and then defend what score they feel they have earned and why. The teacher can conference with the student to discuss any discrepancies between how the teacher scores them and how they score themselves. Students could also answer reflective questions (a *script* form of self-assessment per Panadero et al., 2016) on what skills they want to work on in their writing for the next essay. For the next essay, the teacher then might have students work on newly targeted skills. If some students might be overwhelmed with a long checklist, they could pick a few skills from the checklist on which to focus.

Self-Assessment in Mathematics

Many forms of learning assessments are used in mathematics. Teachers use traditional forms of formative and summative assessments to measure if students are understanding computational processes. For advanced learners to show what they understand and are able to do, it is important to use authentic assessments such as performance-based tasks (VanTassel-Baska & Little, 2017) which often are open-ended. An example might be:

> *You have been asked to organize a room for a meeting of 5 groups of people, ranging from 6-10 per group. You have the option of using round tables that accommodate 8 people per table or rectangular tables around which 6 may be seated. You also need to set up a main area for group leaders to report out their findings and an exhibition for materials to be displayed. Create a diagrammatic model of the room and provide a rationale for your choices.*

Some students get caught up in having the "right" answer and do not like to show any work because they can solve problems in their head. It can be helpful to have students use self-assessment to evaluate how they are solving problems and to be able to

Name: ______________________________________ Date: __________________

Paper Title: ____

Rubric for Informative and Explanatory Writing

Directions: Look at the rubric below. Read through your paper and score yourself on each of the categories below. Circle the description for each category that best describes your paper. Thenput a justification in the column on the right for why you believe that is your score.

	4	**3**	**2**	**1**	**Reason for My Score**
Introduction	The writer introduces the topic with a strong "hook" and previews what is in the body of the paper with a well-developed thesis.	The writer introduces a hook and includes a thesis, but the thesis could preview the topics a little more clearly.	The writer introduces the topic but does not include a thesis to preview the topics of the body of the paper.	The writer introduces the topic with minimal detail and includes little to no thesis.	
Organization	The writer uses a clear organization strategy to organize ideas, concepts, and information. Where appropriate, headings and graphics are used to aid comprehension.	The writer uses a clear organization strategy to organize ideas, concepts, and information but does not include the use of headings or graphics.	The writer uses limited tools to organize the paper.	The writer has used little to no clear organization tools.	
Topic Development	The writer develops the topics with relevant explanations that include definitions, details, and quotations. When quotations are used, thoughtful analysis is also included to explain the significance of the quote.	The writer develops the topics with relevant explanations that include definitions and some details. Quotes are included but need analysis. The process for selecting the best quotes is not clear.	The writer provides some explanation and definitions but may have chosen quotes that were not the best selection for the topic to support or did not use quotes at all for evidence.	The writer develops the topic with minimal explanations and details.	

Transitions and Flow of Paper	Transition words and phrases are used to help the paper flow well between ideas and to create a cohesion at both the beginning and ends of paragraphs.	Transition words and phrases are used to help the paper flow well between ideas and to create some cohesion within and across paragraphs.	The writer attempts to use transitions, but they are either repetitive or inappropriate, resulting in lack of coherence.	The writer did not use transitions in the paper to create a cohesive flow of ideas.	
Language and Style	Domain-specific vocabulary is used to explain concepts discussed. The writer maintains a formal tone throughout the paper.	Domain-specific vocabulary is used to explain concepts discussed. The writer maintains a formal tone throughout the paper most of the time.	Domain-specific vocabulary is used to some extent, but the writer could have used a more formal tone in the paper.	The writer uses very little domain-specific vocabulary and could have used a formal tone in the paper.	
Conventions	The writer shows few to no errors in mechanics such as grammar and spelling. Proper formatting guidelines are followed (i.e., MLA format).	The writer shows some errors in mechanics such as grammar and spelling, but they are not numerous. Proper formatting guidelines are followed (i.e., MLA format).	The writer shows numerous errors in mechanics such as grammar and spelling. Proper formatting guidelines were not followed (i.e., MLA format).	The writer shows many errors in mechanics such as grammar and spelling and did not follow formatting guidelines (i.e., MLA format).	

FIGURE 12.2 Sample writing rubric

explain it (even if they do it a different way). Ask students to create and record in a math journal an explanation of their reasoning when solving problems and to analyze the types of errors they are making.

Students given the opportunity for acceleration can be asked to self-assess, using a checklist to see if they are meeting their goals to work ahead or at a faster pace. Students can also be coached to self-assess through self-grading. The teacher could keep an answer key where students could grade themselves on daily learning practice assignments to see if they are solving problems correctly. This approach allows students to get more immediate feedback (versus waiting for a teacher to grade). Students could keep a record in a math journal of what corrections they are making as they self-grade. They then could conference with the teacher on their learning issues.

When designing curriculum for advanced students, assessment plays an important role in offering both enrichment and acceleration. One approach to differentiating units is to pre-assess students first. Students could either be accelerated to a new unit if they have mastered the concepts in the unit, or they could be offered enrichment

Name: ______________________________ Date: ______________

Rubric for a Math Project

Directions: After completing your math project, rate yourself on the rubric below. Circle the number for the level you feel correctly reflects your achievement on your project. Answer any questions it asks in the "Scoring Criteria" column.

Level of Achievement	Scoring Criteria
4 Exemplary Achievement	I have gone above and beyond what was required in the following ways: These things are extra-special about my product:
3 Proficient (Expected) Achievement	• I have completed my product on time. • I have included my Process Log. • I have done my best work. • I have followed the directions, and my work is neat. • My product shows that I understand the big ideas and skills of the unit and can apply them.
2 Limited Achievement	I have tried to follow directions, work neatly, and show what I know, but I know this is not my best effort. My product could have been better in the following ways:
1 Minimal Achievement	This product did not get much attention or effort from me. There are many ways I could improve this product. Some of them are listed below:
0 Not Able to Be Scored	• I did not complete the assignment.

FIGURE 12.3 Sample elementary math project rubric

projects to dive deeper into a concept. Tomlinson and Eidson (2003) offer practical examples of differentiated elementary math units that use learner contracts, menus, and rubrics to differentiate for students. Students can practice self-assessment throughout this type of unit. Contracts work like a self-marking checklist to have students agree to what they are supposed to complete and to monitor their progress. Students may use rubrics to self-assess their performance on the unit. Another approach is to offer all students an opportunity to do a project in math, but offer differentiation in process, product, or content. To differentiate product options, a menu is useful to provide choices. If needed, different rubrics can be created for different products, or the same rubric can be tailored to fit different projects. Having students co-create the rubric is also helpful. See Figure 12.3 for an example of such a rubric adapted from Tomlinson and Eidson (2003). This rubric asks students to reflect on their psychosocial skills during the unit, their mathematical process and thinking skills, and their final product.

Challenges in Using Student Self-Assessment

Students need training on how to self-assess so they do not underestimate or overestimate their performance and/or product. Teachers also need instruction in both the use of curriculum approaches that are student-centered as well as in student assessment of their own learning.

An important component of training students to use self-assessment is using these techniques frequently in the classroom. Teachers should model using think-aloud protocols on how to complete each method of self-assessment they use in the classroom. Indeed, modeling has been shown to be highly effective in helping students to learn how to accurately self-assess (Kostens et al., 2012). It is also helpful for teachers to have individual conversations with students on how accurate their self-rating appears to be. As mentioned earlier in this chapter, self-assessment is even more effective with feedback from the teacher (Sitzmann et al., 2010). Multiple self-assessment techniques can be used simultaneously. Students could use a checklist such as in Figure 12.1 to evaluate if they included all the necessary components of an essay or project. Before turning in a final project, they could self-assess, using a rubric with specific criteria, such as found in Figure 12.2, that the teacher will also use to evaluate the project.

Teachers themselves may need training, support, and examples of student self-assessment methods to be able to confidently introduce them to students. In addition, teachers may need training to help choose which self-assessment technique is most aligned to specific learning objectives, as well as to help develop rubrics and other self-assessment materials that are differentiated. Self-assessment methods are particularly well-suited to student-centered curriculum approaches that encourage critical thinking, inquiry, and creativity. Thus, teachers may also need training and support to implement these more student-centered curricular approaches.

Recommendations for Coordinators and Directors of Gifted Education Programs

Coordinators and directors of gifted education programs should consider the important role self-assessment can play in academic talent development for gifted learners. Self-assessment can enhance learning and performance, as well as increase the psychosocial skills of self-regulation and self-efficacy, both of which further enhance learning and performance, as well as metacognition. Designing learning opportunities that intentionally incorporate self-assessment can help students develop those skills.

- Gifted coordinators should review gifted programs and curriculum to evaluate where there is opportunity to build in self-assessment techniques. The form of self-assessment can be chosen based on the type of skill that teachers and coordinators are wanting to develop in that learning opportunity. Self-assessment is a skill that benefits all students but can play an especially important role in the development of academically advanced students and be a guiding force in the use of differentiation to support those students. Gifted coordinators should work with teachers to help them understand the different self-assessment techniques so they can determine the best methods for specific tasks.
- Gifted coordinators should consider the suggestions listed above for how self-assessment can be used in programs for gifted learners. Pre-assessment plays an essential role in making sure students are placed in the appropriately advanced courses as does the use of self-pacing when they are placed in advanced curriculum. Learning

contracts, self-marking checklists, and rubrics can all be used as self-assessment tools in conjunction with both acceleration and enrichment experiences.

- Student self-assessment is likely to be more effective with training, and this extends to the training of the teachers regarding how and when to use it. Self-assessments can also be differentiated to match curricular adaptations and can be varied for process, product, and content.
- Gifted coordinators can work with teachers to create checklists and differentiated rubrics. In addition to differentiated rubrics and checklists, portfolios are another component that could be added to gifted self-assessment approaches. Portfolios provide an avenue for students to build a collection of the work that reflects their creativity and deeper learning. Advanced students could also work with gifted coordinators, teachers, and counselors to set goals for their academic advancement and to monitor how they are progressing with their goals.
- Student self-assessment can be used as an important tool to evaluate the efficacy of the programs in which gifted learners are being served. Coordinators and directors of gifted education programs can use student self-assessment data to evaluate the impact of gifted programming on student learning and performance while also providing "student voice" in the process.

Conclusion

Student self-assessment is just one component of assessment of learning, but it can be used in a variety of ways to enhance student understanding of their own progress in the learning process as seen in this chapter. To build strong programs that differentiate for gifted learners and allow for acceleration and enrichment, coaching students to use self-assessment tools and techniques will play a vital role in building independent and self-regulated learners.

References

Alonso-Tapia, J., & Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, *33*(3), 385–397. https://doi.org/10. 1174/021037010792215145

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287–301, 320. https://doi.org/10.3200/JOER.102.4.287-302

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Boud, D., & Brew, A. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education*, *18*, 130–135.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, *18*(5), 529–549. https://doi.org/10.1007/BF00138746

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Sage.

Burns, R. A., Crisp, D. A., & Burns, R. B. (2020). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the cross-lagged panel and random-intercept cross-lagged panel frameworks. *British Journal of Educational Psychology*, *90*(1), 77–91. https://doi.org/10.1111/bjep.12265

Common Core State Standards Initiative. (2021). *English Language Arts Standards Writing Grade 7*. www.corestandards.org/ELA-Literacy/W/7/

Dixson, D. D., Worrell, F. C., Olszewski-Kubilius, P., & Subotnik, R. F. (2016). Beyond perceived ability: The contribution of psychosocial factors to academic performance. *Annals of the New York Academy of Sciences*, *1377*(1), 67–77. https://doi.org/10.1111/nyas.13210

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395–430. https://doi.org/10.3102/00346543059004395

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, *53*, 69–76. https://doi.org/10.1016/j.stueduc.2017.03.003

González-Betancor, S. M., Bolívar-Cruz, A., & Verano-Tacoronte, D. (2019). Self-assessment accuracy in higher education: The influence of gender and performance of university students. *Active Learning in Higher Education*, *20*(2), 101–114. https://doi.org/10.1177/1469787417735604

Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2016). The relationship between self-efficacy and academic achievement: A 5-year panel analysis. *The Asia-Pacific Education Researcher*, *25*(1), 89–98. https://doi.org/10.1007/s40299-015-0236-3

Kanevsky, L. (2011). Deferential differentiation: What types of differentiation do students want? *Gifted Child Quarterly*, *55*(4), 279–299. https://doi.org/10.1177/0016986211422098

Kostens, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, *22*(2), 121–132. https://doi.org/10.1016/j.learninstruc.2011.08.004

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lent, R. W., & Brown, S. D. (Eds.). (2017). Social cognitive career theory in a diverse world [Special issue]. *Journal of Career Assessment*, *25*(1). https://doi.org/10.1177/1069072716657811

Lew, M., Alwis, W. A. M., & Schmidt, H. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment and Evaluation in Higher Education*, *(35)*2, 135–156. https://doi.org/10.1080/02602930802687737

Machado, L., & Yoshinaga, C. (2018). Self-assessment accuracy, overconfidence and student performance. *The New Educational Review*, *53*(3), 85–94. https://doi.org/10.15804/tner.2018.53.3.07

Oppong, E., Shore, B. M., & Muis, K. R. (2019). Clarifying the connections among giftedness, metacognition, self-regulation, and self-regulated learning: Implications for theory and practice. *Gifted Child Quarterly*, *63*(2), 102–119. https://doi.org/10.1177/0016986218814008

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, *39*(3), 125–132. https://doi.org/10.1016/j.stueduc.2013.04.001

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org.10.1016/j.edurev.2017.08.004

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806–813. https://doi.org/10.1016/j.lindif.2012.04.007

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/bf00117714

Siegle, D. (2002). Creating a living portfolio: documenting student growth with electronic portfolios. *Gifted Child Today*, 25(3), 60. https://doi-org./10.4219/gct-2002-70

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning and Education*, *9*(2), 169–191. https://doi.org/10.5465/amle.9.2.zqr169

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (Eds.). (2018). *The psychology of high performance: Developing human potential into domain-specific talent*. American Psychological Association.

Snyder, K. E., Nietfeld, J. L., & Linnenbrink-Garcia, L. (2011). Giftedness and metacognition: A short-term longitudinal investigation of metacognitive monitoring in the classroom. *Gifted Child Quarterly*, *55*(3), 181–193. https://doi.org/10.1177/0016986211412769

Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, *15*(2), 199–209. https://doi.org/10.1080/13562511003620027

Tomlinson, C. A., & Eidson, C. C. (2003). *Differentiation in practice: A resource guide for differentiating curriculum, grades K-5*. Association for Supervision and Curriculum Development.

Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. Association for Supervision and Curriculum Development.

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards Vol. 1* (pp. 55–87). Springer.

VanTassel-Baska, J., & Brown, E. F. (2007). Toward best practice: An analysis of the efficacy of curriculum models in gifted education. *Gifted Child Quarterly*, *51*(4), 342–358. https://doi.org/10.1177/0016986207306323

VanTassel-Baska, J., & Little, C. (Eds.). (2017). *Content-based curriculum for high ability learners* (3rd ed.). Prufrock Press.

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. Routledge.

CHAPTER

13

The Use and Value of State Assessments of Learning

Cecelia A. Boswell, Cheryll M. Adams, and Mary M. Christopher

Introduction

Assessment is crucial to monitoring the progress and evaluating the growth of gifted students as a result of the programs and services provided for them (Callahan et al., 2017; Cao et al., 2017; Robbins, 2019; Ryser & Rambo-Hernandez, 2014; VanTassel-Baska, 2006, 2019). Schools must use data to show growth for all students, even gifted learners, as part of the accountability movement; thus, schools should examine new ways to accomplish this important task (Council for Exceptional Children, The Association for the Gifted 2012; McCoach et al., 2013; Rambo-Hernandez & McCoach, 2015; Ryser & Rambo-Hernandez, 2014).

The Use of State Assessments

State-level data on student achievement in the United States come from two primary sources—state tests and the National Assessment of Educational Progress (NAEP). State assessments that report academic growth vary as widely as the states themselves and their approaches to gifted programming and services. The value of these assessments differs according to fidelity of implementation and their alignment with policies and practices at the school level. Although a few states use nationally normed standardized achievement measures, (e.g., NWEA MAP), most use assessments aligned with their own state standards in the core academic subjects, specific to each grade level being assessed. They may be useful for monitoring progress among general education students, planning interventions, and evaluating the effects of programming for these students.

Problems in the Use of State Assessments to Document Learning for Gifted Students

For gifted and advanced students, however, these assessments are problematic. One problem, related to grade level standards, is the paucity of items that are challenging enough to differentiate gifted students' level of performance from other typical students

DOI: 10.4324/9781003285991-18

who have many questions that are challenging for them, thus introducing greater than expected measurement error. In common parlance, this means that the tests are not difficult enough to discern the actual learning accrued for gifted students in the area being tested (i.e., ceiling effect). While most states have cut-off scores that define various levels of achievement at specific grade levels (e.g., Below Expectations, Proficient, Exceeds Expectations), gifted students' true achievement level cannot be measured on these tests because there is not enough ceiling for students to show growth (Lakin & Rambo-Hernandez, 2019; Lohman & Korb, 2006).

Another problem relates to how the state data are used or not used in respect to gifted student performance. States do not fully disaggregate the scores on their statewide assessments by subgroup, proficiency level, and grade level, and most states do not disaggregate the scores of gifted learners on these assessments. Furthermore, the definitions of proficient and advanced vary from state to state as well as the rigor of the content being tested, contributing to the difficulty in finding these "excellence gaps" at the state level (Plucker, et al., 2010, Plucker & Peters, 2016).

The Use of National Assessment Tools

The NAEP, which is administered every two years in reading and math and less often in other subjects, reports national results at grades 4, 8, and 12 and state-by-state results at grades 4 and 9. The content on this test is not aligned with any state's standards but rather frameworks developed by a National Assessment Governing Board appointed by the U. S. Secretary of Education. Results from NAEP indicate a closing of the achievement gap at the basic and proficient levels, but the gaps at the highest level of achievement do not appear to be shrinking (Plucker et al., 2010; Plucker & Peters, 2016). The most recent NAEP data (National Center for Education Statistics, 2019) indicate that the percentages of students scoring at or above NAEP Proficient has remained unchanged in mathematics and is lower in reading than in 2017, supporting the earlier conclusions of Plucker and his colleagues (2010, 2016) that while gaps at the Below Basic, Basic, and Proficient levels are decreasing, the percent of students scoring at or above Proficient is not increasing. Thus, most students are graduating from high school without advanced understanding in the content areas.

Finally, although national assessments can provide snapshots of how well our states and nation are performing in comparison to others, they cannot examine individual students' scores over time because they are administered periodically to representative samples of students in selected schools rather than annually to all states.

The Use of Growth Models

Growth modeling is a possible procedure for tracking students' academic progress over a period of time (McCoach et al., 2013; Rambo-Hernandez & McCoach, 2015; Ryser & Rambo-Hernandez, 2014). McCoach et al. (2013) underscore that statistically sound growth models must have at least three observations using assessments that are psychometrically sound and comparable across time. According to Castellano and Ho (2013), value-added growth models examine information about what affects growth, such as particular teachers or programs, depending on the factors selected. Using these models, schools can report that the growth made beyond what was controlled for (e.g., age, prior achievement, ethnicity, socioeconomic status) was due to the value added by the program. Growth models in general can be a fairer assessment of accountability because they involve scores across time rather than a single point (Robbins, 2019). The difficulty with tracking students' academic progress through growth modeling, however, occurs when using standardized tests and instruments that have low ceilings because, as noted earlier, gifted learners generally score at the top of these measures. This makes it almost impossible to demonstrate that gifted learners have progressed commensurate with their ability because there is little room for growth on such test scales.

In summary, states have developed state-wide assessments for accountability purposes under Every Student Succeeds Act (ESSA, 2015). These assessments generally do not measure gifted learner progress, which is usually at or above proficiency and often lack enough challenging questions for them to show growth (Ryser & Rambo-Hernandez, 2014).

Approaches to Address the Problems of State Assessment Use with Gifted Students

Above-level testing (e.g., giving a third grader the fourth-grade test) can address the two issues mentioned above and provide a better assessment of what the student actually knows, understands, and can do. Because the Every Student Succeeds Act (ESSA, 2015) does allow above-level assessment, some states are using computer adaptive assessments as part of their mandated achievement testing. However, we could find no instances of states allowing above-level testing on grade-level state-wide assessments, thus making it difficult to use these assessments to show adequate yearly progress for gifted learners.

Using state assessments as one of several indicators of achievement for gifted students in a content area also increases their value. When state assessments are used alone, they do not generally provide the information necessary to determine the effectiveness of gifted programs and services. However, when used as one of several tools in a program evaluation, they can reveal areas of concern or commendation (Speirs Neumeister & Burney, 2018, 2019). Interviews with stakeholders, document reviews, classroom observations, curriculum audits, and surveys are other tools that can be used to conduct an in-depth program evaluation. Other assessment indicators useful in programs for the gifted have traditionally included performance-based measures such as writing assessments and project work and portfolios that illustrate students' perception of their own learning (see Johnsen, 2012; Johnsen et al., in press; VanTassel-Baska, 2008).

Petrilli (2016) offers four elements that must be present if learners above the proficient level are going to matter in overall assessment approaches:

- Give schools extra credit for getting students to the advanced levels on state tests,
- Use a growth model, such as value-added, that looks at the progress of all students at all achievement levels, not just overall proficiency,
- Make growth matter the most when determining summative school grades or ratings for entry to programs for the gifted, and
- Include gifted students as a subgroup when reporting growth.

He cautions that continuing to look at the lowest performing groups has the unintended consequence of making low achievers a higher priority than high achievers; in high poverty schools, this hurts high-achieving, low-income students the most.

Current Status of States that Include Assessment of Learning

One goal of this chapter is to share data about those states that offer data on the annual academic growth of their gifted students. NAGC offered an overview of the states that do provide information relative to assessments for growth in gifted students, states that do not provide the information, and those that have no policy, ergo, no information. The State of the States Report (Rinn et al., 2020) noted that only 11 of the 51 possible respondents provided a link to their most current annual report. Eleven states responded that they did offer gifted students as a subgroup for state accountability reporting and provided the information on their website. These 11 states indicated a policy for assessment, but the reports were limited in respect to overall data. Neither disaggregated data for the gifted, nor information relative to criteria for the development of assessments was available. Only Ohio provided criteria for assessment policy and practice.

After reviewing the NAGC report, we attempted to gather information about the assessment of gifted students from state departments of education websites, but, in some cases, information was accessed only by local school districts and/or parents in that state. We therefore had conversations with state directors and other state personnel knowledgeable about assessments of their gifted students' learning and reviewed states' ESSA reports and other advanced assessment information to determine the current status of each state's assessment of gifted learners that guide the states in developing their assessments for learning progress (see Table 13.1 for definitions).

As state personnel change, the institutional knowledge is often lost. The rationale used in the development of assessments was unknown to current directors of Gifted or Advanced Academic departments. Hence, we could not establish any criteria to use to evaluate the assessments used by states to judge gifted learner growth. Therefore, we organized this chapter according to insights gained from the information gathered from state education agencies and their websites.

After reviewing states' websites for gifted or high ability learners, only 14 were found that provided laws or standards for assessing gifted students (see Table 13.2). While their laws or standards called for alternative measures of student progress or use of standardized

TABLE 13.1 State assessment information related to gifted student learning growth

Descriptor	Definition
Gifted Assessments in State Law or State Standards	States whose website offers information about assessment for describing compliance with state law or state standards
State Assessments Available by Gifted Subgroup	States that list assessment information for gifted as a subgroup.
Advanced Assessment and Above Level Progress	States that provide assessment information that includes AP, IB, Dual Credit and students who are above grade level,
Inclusion in ESSA	States that provide assessment information as a part of their ESSA documents
Alternative Assessments	States that provide assessment information in other ways than state-wide or district-level testing
Assessments Used or Future Planning	States that indicate that assessment information is used beyond just reporting

TABLE 13.2 Current status of assessments of gifted student learning growth (N=17)

State	Include Gifted Assessments in State Law or State Standards	Make State Assessments available by gifted subgroup	Report Above level progress	Report Advanced Assessment results (IB/AP/Dual Credit)	Use Alternative Assessments	Inclusion in ESSA	Use Assessments for future planning
Arkansas	X						
California	X						
Colorado	X	X			X	X	
Delaware	X						X
Florida	X						
Illinois	X						
Indiana	X	X	X	X		X	
Kentucky	X						
Louisiana	X	X					
Maryland		X					
Nevada				X		X	
New Mexico	X						
North Carolina	X			X			
Ohio	X	X			X		
Oregon	X						
Texas					X		
Washington	X						

tests to report academic growth, either the data were not available or only found in an individual school search. Some states requested the information but recognized that local districts can decide if relevant data are sent to the state or kept secure because of privacy issues. There are instances in which state law requested gifted program assessment but did not require separate assessment data for gifted students within the programs.

Five states said that they report student growth for gifted students as a subgroup and offered the information in a variety of ways. These states established performance expectations with qualitative and quantitative evidence and provided data by school and grade level. At the same time, they often limited access to the district and its parents or included gifted in their value-added model.

Only three of the 17 states disaggregated the data for gifted students' assessments in advanced classes. Yet even these states did not offer the information to the public.

Three states included gifted in the narrative of their ESSA plans but only as part of a list with other underrepresented populations. Analyses of ESSA plans revealed only three states whose information addressed gifted learners as an integral part of assessments for academic growth. These three states provided advanced assessment reports for gifted as a subgroup on their websites.

There were three states whose assessments went beyond state-wide or district-level testing. One state, Delaware (Delaware Department of Education, 2014), used the assessments for future planning. The three states that offer alternative assessments did so through differing lenses. The section below, Alternative Growth Assessments, describes their approaches and the approach used for future planning. Information that is not available to the public through the states' websites relates to local districts that maintain the information. Of the nine states that offer some information, six referred the reader to the individual schools to access their data. Two states said that they simply collect the information. One state reported that they offer growth information, but required data only related to identification of gifted students. One state included information about twice exceptional learners through special education sources, and one revealed that the data will become available after resolution of the impact of COVID-19.

States also use different terms to describe the gifted population and different requirements for reporting the information. For example**,** Indiana uses the term High Ability (HA) rather than gifted when referring to this population (Indiana's Learning Evaluations Assessment Network (ILEARN) Program, n.d.). Their state code includes the following concerning documentation of learning growth:

> **(b)** The department shall disaggregate from the total results of the statewide assessment test results for a school corporation the percentage of students in each school and each grade in the school corporation that are identified as high ability students (as defined by IC 20-36-1-3) by the school corporation who also achieved a score in the highest performance level designated for the statewide assessment (Indiana's Learning Evaluations Assessment Readiness Network [ILEARN] Program, n.d., Annual performance reports; disaggregation of test results for high ability students)
>
> However, this disaggregation is not required in a case in which the results would reveal personally identifiable information about an individual student under the federal Family Education Rights and Privacy Act.

Another example is North Carolina that sets program standards for their intellectually gifted program. Standard 6, Part d, addresses program accountability, stating each

LEA, "maintains, analyzes, and shares student achievement, student growth, and annual drop-out data for AIS students" (North Carolina Academically or Intellectually Gifted Program Standards, 2018, p. 6). Even though each Local Education Agency (LEA) provides data for their district, no state composite can be obtained for gifted learners as a subgroup.

An integral part of the accountability movement involves determining student growth through assessment data (Ryser & Rambo-Hernandez, 2014), but state assessment of gifted student academic growth varies in approach, quality, and fidelity to state laws and standards. None meet the criteria established by Petrilli's (2016) four elements that provide a clear picture of above-level students' achievement.

Alternative Growth Assessments

The National Association for Gifted Children offers an Administrator's Toolbox that provides insight into accountability measures appropriate to gifted learners. Because of ceilings on many forms of assessments, the expected growth may not be observed. Educators who are aware of the limitations of their selected or mandated assessments can plan for alternative tests and other types of assessments. NAGC recommends, "Items that assess critical thinking and not just knowledge and comprehension are needed" (National Association for Gifted Children, n.d.).

State ESSA reports encourage effectiveness of gifted learners' programs in relation to growth assessments. Currently, assessment for growth in gifted learners is required or addressed specifically in the reports in a few states. This information can provide leadership for states that are looking to meet the ESSA requirement of presenting such information.

Colorado Growth Assessment: Unified Improvement Plan

Colorado is among the few states that attempts to measure academic growth. The Exceptional Children's Educational Act (ECEA), authorized in 2014, outlines the guidelines for gifted education programs and specifically includes a section on evaluation and accountability. The procedures include the development of a biannual Unified Improvement Plan (UIP) "to align efforts to improve gifted student achievement and growth" (Colorado Department of Education, 2018). Administrative Units (AU) integrate performance data into an electronic UIP form that requires them to analyze gifted students' performance in order to set relevant goals. Districts then develop targets related to areas of growth for their gifted learners. One limitation on measurement of learning progress particular to gifted students results from the state's focus on local control. The Colorado Department of Education, provided the following information about measurement of student learning.

> Colorado does monitor student growth for gifted learners, using median student growth percentiles as well as through academic performance on our state assessments. Our gifted leads at the local level look at growth in ranges from low, typical, to high for the gifted learners in their system. This approach allows our gifted leader to look at how gifted learners are preforming compared to their gifted peers as well as traditional learners within their system. Additionally, we have mandated Advanced Learning Plans in which every identified gifted learner has both academic and

> affective goals which are monitored annually. These goals are a targeted way to look at individual student growth year to year (personal communication, April 8, 2021).

The state offers insight into the development of their state assessments focused on gifted learner's growth. The state focuses on gifted students and their education in Section 6, Supporting All Students. "The strategies and uses of funds must be designed to ensure that all children have a significant opportunity to meet challenging State academic standards and career and technical standards" (Colorado Department of Education, 2018, p. 108).

Colorado also establishes performance expectations. The expectations define in specific terms what growth should include and a means for qualitative information as well as numerical data (Colorado Department of Education, 2020). Each district may access scores to measure student growth for their gifted students in evidence-based reading and writing, math, PSAT/SAT.

Finally, Colorado offers growth assessment through Advanced Learning Plans. While these plans rely on test scores, they are a unique approach to identifying learning needs of gifted students at the local level (R. McKinney, personal communication, April 8, 2021).

Research and evidence that considers supports for exceptional students, including students who are gifted and exhibit a disability, emphasizes the importance of recognizing exceptional potential in early years and developing talents and abilities over time in a purposeful manner. Through policy and state/local resources, Colorado has committed to discover exceptional potential in every student population so that all student groups, especially underrepresented populations, have access to gifted programming. Multiple pathways into gifted programming result in long-term planning and services for appropriate coursework and post-secondary outcomes.

Ohio Growth Assessment: The Gifted Indicator

According to Ohio's ESSA document, it is one of the only states in the nation to rate schools and districts on the performance of and opportunities provided to its gifted students (Ohio Department of Education, 2018). The gifted subgroup is defined in law as students identified as gifted in superior cognitive ability and specific academic ability (Ohio Revised Code, 2019). Ohio evaluates student growth by using growth information to determine how programs and services benefit the gifted students. This method analyzes student growth information to establish the level of services and their impact on student academic growth. Beginning with the Ohio Report Card for the 2014–2015 school year,

> The performance indicators shall include an indicator that reflects the level of services provided to, and the performance of, students identified as gifted under Chapter 3324 of the Revised Code. The indicator shall include the performance of students identified as gifted on state assessments and value-added growth measure disaggregated for students identified as gifted. (Ohio Revised Code, 2015, p. 1)

This Gifted Indicator (Ohio Department of Education, 2020a) identifies the number of a district's or school's students who are identified as gifted and how many of them are receiving gifted services. Additionally, it shows how gifted students are performing academically on state-wide assessments, reflects a district's Met/Not Met indicators for state assessment of gifted services, and appears on the Ohio School Report Card.

The Gifted Performance Index (PI) is determined by calculating students' scores on their state-wide assessments. Scores are weighted by the level of proficiency (Limited, Basic, Proficient, Accelerated, Advanced, and Advanced Plus) and multiplied by the number of scores at each level. The scores are added to produce a PI between 0 and 120. The Gifted PI must be at least 117 to meet the requirements for this portion of the Gifted Indicator (Ohio Department of Education, 2020a).

The Gifted Value-Added is a graded measure on the Ohio State Report Card, indicating the impact the district or building had on the yearly growth of the students. Gifted students' scores on designated state assessments are calculated to determine a growth estimate which is transformed into a growth index. A letter grade is assigned based on the growth index. Currently, a grade of A is earned if the growth index is greater than or equal to +1 and a C is less than or equal to −1 but greater than or equal to −2. Districts must receive a grade of C or better to meet the requirements for this part of the Gifted Indicator (Ohio Department of Education, 2020b).

In summary, Ohio disaggregates achievement scores of four different subgroups, one of which is the gifted subgroup, and they examine how gifted students are identified and served. All of these data points are quantified to produce a Gifted Indicator score that reflects how well a school or district is meeting the expectations for gifted performance and progress.

Texas Growth Assessment: Texas Performance Standards Project

Texas educators of the gifted/talented did not want a minimum standard as set in the No Child Left Behind Act to be applied to gifted learners. As they were developing the Texas Performance Standards in 2000 (2000–2006, Texas Performance Standards Project, Executive Summary), this state-wide committee considered that the requirements of the State Goal (2009) for the Gifted/Talented could be met through a performance-based project.

In addition to the State Goal, the Texas State Plan for the Education of the Gifted/Talented (2019) (State Plan) states in Section 3: Curriculum and Instruction that districts meet the needs of gifted/talented students by modifying the depth, complexity, and pacing of the curriculum and instruction ordinarily provided by the school. The Accountability piece of this section offers that an array of appropriately challenging learning experiences is provided in core areas and that students may purse areas of interest within a continuum of services that lead to development of advanced-level products.

To meet the state law and State Plan, the Texas Performance Standards Project (TPSP) provides open-ended units of study that include a Texas Essential Knowledge and Skills (TEKS) alignment guide, adaptable activities, and other resources. The structure and content of the tasks provide the following:

- Wide variety of choices for student learning
- Flexibility to pursue topics of student interest
- Real-life research experiences
- Focus on a high-quality product and presentation (TPSP, n.d., sec. 2)

The assessment section of students' individual TPSP for grades 1–10 defines the six domains applied to students' projects. They cover content knowledge and skills, analysis and synthesis, multiple perspectives, research, communication, and presentation of learning. Grades 9–12 add the four domains of ethics/unanswered questions, methodology and use of resources, relevance and significance, and professional quality.

Scoring dimensions are intended for sharing with students at the beginning of their tasks or independent projects. When students access scoring tools prior to and during their work on projects, they self-evaluate and discuss the strengths and areas of need with their teacher or mentor who gives a final assessment. Use of the rubrics made appropriate for each grade level provides qualitative assessments for students' learning as they progress toward the standard set in the Texas State Goal for the Gifted/Talented.

Recommendations

Based on a review of available data from the State of States Report (Rinn et al., 2020), state websites, and personal communications, we determined that some changes in the use of state and national data would support effective assessment of gifted learners. Although most states required annual assessments of state standards for all students, a clear focus on gifted learners remains essential to the academic growth of this special population.

1. **State Departments of Education need to develop clear policies and tools for annual reports relating to assessment of gifted and talented K-12 students. Educational policy drives the expectation for appropriate services.** Previous reports (Assouline et al., 2015; Colangelo et al., 2004; National Commission on Excellence in Education, 1983; U.S. Department of Education, 1993) outlined the need for a focus on American's best and brightest students, but inconsistent policy has remained in most states. Our review determined a continued lack of attention to policy and to enactment even when policy is in place.
2. **Gifted and talented students need to be included as a subgroup in district and state accountability measures.** Most states do not report on this special population as a separate group. Making gifted and talented students a trackable subgroup emphasizes the relevancy of their learning to the overall accountability of districts.
3. **All states' rules need to follow the expectations of high achieving students outlined in ESSA.** The indicator of academic achievement required by ESSA provides an opportunity for schools to earn extra credit for students reaching advanced academic levels (Petrilli, 2016). In ESSA, LEAs and states must "collect, disaggregate, and report their student achievement data at each achievement level" (Every Student Succeeds Act, 2015, p. 2). If states begin to implement ESSA accountability requirements, they will move to a larger focus on high achievement rather than proficiency, therefore providing expectation of a quality education for *all* students regardless of race, ethnicity, socio-economic status, or language proficiency.
4. **State rules need to include growth assessment for all achievement levels of accountability measures.** Gifted students who are appropriately challenged may demonstrate academic growth of 18–21 months (or more) in a year's time. One issue related to the use of state assessments is that they have low ceilings, making

growth assessments of academic progress limited for gifted students. Rather than focusing on growth toward proficiency on state standards, state assessments need to track students' advanced achievement over time. A growth model would measure academic success based on individual student gains and improvement beyond grade-level (Council for Exceptional Children, The Association for the Gifted, 2012).

5. **Accountability requirements need to add the use of alternative assessments that measure higher level outcomes (e.g., scientific research, critical thinking, creative thinking).** Learner outcomes for advanced and gifted learners create a focus not only on content but methodology or processes used within specific disciplines and professional settings. These processes support in-depth learning that exhibits complex and abstract thinking (Callahan, 2009; Kaplan, 2009; Reis & Renzulli, 2009; Texas Education Agency, 2019). The inclusion of alternative assessments in state accountability models will support high-level expectations for gifted and talented learners.

Conclusion

While it remains difficult to assess student growth in gifted learners, state assessment standards and procedures cannot ignore this important population and focus on minimal competency rather than excellence. Implementing the outlined recommendations will support assessment of academic growth in gifted learners. Measuring student achievement over time creates a dynamic view of school effectiveness that closes the excellence gap, supports the addition of excellence to the minimum competency standard, and addresses accountability in gifted education (McCoach, Rambo, & Welsh, 2013; Plucker, Burroughs, & Song, 2010).

References

Assouline, S. G., Colangelo, N., & VanTassel-Baska, J. (2015). *A nation empowered: Evidence trump the excuses holding back America's brightest students (Vol. 1).* The Connie Belin & Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Callahan, C. M., Moon, T. R., & Oh, S. (2017). Describing the status of programs for the gifted: A call to action. *Journal for the Education of the Gifted*, *40*(1), 20–49. https://doi.org/10.1177/0162353216686215

Callahan, C. M., (2009). Evaluation for decision-making: The practitioner's guide to program evaluation. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted* (pp. 119–142). Prufrock Press.

Cao, T. H., Jung, J. Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics*, *28*(3), 163–203. https://doi.org/10.1177%2F1932202X17714572

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Washington. DC: Council of Chief State School Officers.

Colangelo, N., Assouline, S. G., & Gross, M. U. M. (2004). *A nation deceived: How schools hold back America's brightest students (Vol. 1).* The Connie Belin & Jacqueline N. Blank Center for Gifted Education and Talent Development. https://www.accelerationinstitute.org/nation_deceived/nd_v1.pdf

Colorado Department of Education. (2018). *Consolidated state plan under the Every Student Succeeds Act (ESSA).* https://education.ky.gov/AA/Pages/default.aspx

Colorado Department of Education (2020). *Program evaluation and accountability guidance*. www.cde.state.co.us/gt/programevaluation-0

Colorado Department of Education. (2021). *Gifted education quality program assessment rubric*. www.cde.state.co.us/gt/giftededqualityprogramassessment

Council for Exceptional Children, The Association for the Gifted. (2012). *Growth in achievement of advanced students*. http://cectag.com/wp-content/uploads/2012/04/Growth-Models.pdf

Delaware Department of Education. (2014). *Gifted or talented education program resource guide*. www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/140/gifted_talented_education_program_resource_guide.pdf.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://congress.gov/114/plaws/publ95/PLAW-114publ195.pdf

Indiana's Learning Evaluations Assessment Network (ILEARN) Program, Indiana Code § 20-32-5.1-16b (n.d.). https://casetext.com/statute/indiana-code/title-20-education/article-32-student-standards-assessments-and-performance/chapter-51-indianas-learning-evaluation-assessment-readiness-network-ilearn-program/section-20-32-51-16-annual-performance-reports-disaggregation-of-test-results-for-high-ability-students

Johnsen, S. (2012). *National standards in gifted education*. Prufrock Press.

Johnsen, S., Dailey, D., & Cotabish, A. (2022). *National standards for gifted education*. Prufrock Press.

Kaplan, S. N. (2009). The grid: A model to construct differentiated curriculum for the gifted. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted* (pp. 235–252). Prufrock Press.

Lakin, J. M., & Rambo-Hernandez, K. (2019). Fidelity of implementation: Understanding why and when programs work. *Gifted Child Today*, 42(4), 205–214. https:doi.org/ 10.1177/1076217519862327

Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted*, *29*(4), 451–484. https://doi.org/10.4219/jeg-2006-245

McCoach, D. B., Rambo, K. E., & Welsh, M. (2013). Assessing the growth of gifted students. *Gifted Child Quarterly*, *57*(1), 56–67. https://doi.org/10.1177/0016986212463873

National Association for Gifted Children. (n.d.). *Administrators Toolkit: Accountability, assessment, and learning growth for gifted students*. www.nagc.org/sites/default/files/administrators/Accountability%20for%20GT%20student%20learning%20(rev).pdf

National Association for Gifted Children. (2019). *The NAGC pre-K to grade 12 gifted programming standards*. www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20Standards%281%29.pdf

National Center for Education Statistics. (2019) *The nation's report card*. http://nces.ed.gov/nationsreportcard/

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. U.S. Government Printing Office.

North Carolina Academically or Intellectually Gifted Program Standards, State Board of Education Policy ACIG-000. (June 2018). www.dpi.nc.gov/media/10010/download

Ohio Department of Education. (2018). *Revised state template for the consolidated state plan*. http://education.ohio.gov/getattachment/Topics/Every-Student-Succeeds-Act-ESSA/OH_ESSA_SUBMISSION.pdf.aspx

Ohio Department of Education. (2020a). *Technical document for the gifted indicator: Gifted indicator description and general business rules*. http://education.ohio.gov/Topics/Data/Report-Card-Resources/Achievement-Component/Gifted-Students-Indicator

Ohio Department of Education. (2020b). *Technical document for 2020 value-added calculations*. https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Progress-Measure/Technical-Documentation-Value-Added-Progress-Dimension.pdf.aspx

Ohio Department of Education. (2020c). *Watch schools and districts*. https://education.ohio.gov/Topics/District-and-School-Continuous-Improvement/Federal-Programs/ESEA-Support-Schools-and-Districts/Watch-Schools

Ohio Revised Code § 3317.40 (2019). https://codes.ohio.gov/ohio-revised-code/section-3317.40#:~:text=Section%203317.40%20%7C%20Accountability%20for%20funding.&text=(1)%20Students%20with%20disabilities%3B,(2)%20Economically%20disadvantaged%20students%3B&text=(4)%20Students%20identified%20as%20gifted,of%20the%20Revised%20Code.

Ohio Revised Code § 3302.02 (2015). https://codes.ohio.gov/assets/laws/revised-code/authenticated/33/3302/3302.02/9-29-2015/3302.02-9-29-2015.pdf

Oregon Department of Education (2018–2019). *An annual report to the legislature on Oregon Public Schools.* www.oregon.gov/ode/schools-and-districts/reportcards/Documents/rptcard2019.pdf

Petrilli, M. J. (2016, March 30).ESSA accountability: Don't forget the high-achievers. *Thomas B. Fordham Institute*. https://fordhaminstitute.org/national/commentary/essa-accountability-dont-forget-high-achievers

Plucker, J. A., Burroughs, N., & Song, R. (2010). *Mind the (other) gap!: The growing excellence gap in K-12 Education*. Center for Evaluation & Education Policy. https://files.eric.ed.gov/fulltext/ED531840.pdf

Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students.* Harvard Education Press.

Rambo-Hernandez, K. E., & McCoach, D. B. (2015). High-achieving and average students' reading growth: Contrasting school and summer trajectories. *The Journal of Educational Research*, *108*(2), 112–129. https://doi.org/10.1080/00220671.2013.850398

Reis, S. M., & Renzulli, J. S. (2009). The schoolwide enrichment model: A focus on strengths and interests. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted* (pp. 323–352). Prufrock Press.

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the States in Gifted Education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. www.nagc.org/2018-2019-state-states-gi

Robbins, J. I. (2019). Gifted program evaluation: Catalyst for change or keeper of the culture. *Gifted Child Today*, *42*(4), 196–204. https://doi.org/10.1177%2F1076217519862326

Ryser, G. R., & Rambo-Hernandez, K. E. (2014). Using growth models to measure school performance: Implications for gifted learners. *Gifted Child Today*, *37*(1), 17–23. https://doi.org/10.1177/1076217513509617

Speirs Neumeister, K., & Burney, V. H. (2018). *Colorado Springs District 11 gifted program evaluation*.Consultants for High Ability. www.d11.org/Page/2552

Speirs Neumeister, K., & Burney, V. H. (2019). *Gifted program evaluation: A handbook for administrators and coordinators* (2nd ed.). Prufrock Press.

Texas Performance Standards Project. (n.d.). *Texas performance standards project.* www.texaspsp.org/how-to-use

Texas Education Agency. (2019). *Texas state plan for the education of gifted/talented students*. https://tea.texas.gov/sites/default/files/GT_State_Plan_2019_1.pdf

U.S. Department of Education. (1993). *National excellence: A case for developing America's talent.* http://nagc.org.442elmp01.blackmesh.com/sites/default/files/key%20reports/National%20Excellence%20%281993%29.pdf

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 school districts: A Clarion call for enhanced gifted program development. *Gifted Child Quarterly*, *50*(3), 339–341. https://doi.org/10.1177/001698620605000302

VanTassel-Baska, J. (2008) Alternative assessment for gifted students. Prufrock Press.

VanTassel-Baska, J. (ed.) (2019) *Gifted Child Today* Special Issue 42, (4).

CHAPTER

14 Assessment of Talent Development Trajectories

Paula Olszewski-Kubilius, Frank C. Worrell, and Rena F. Subotnik

Assessment of Talent Development Trajectories

Historically, giftedness has been viewed as an inherent quality of an individual, largely conceptualized as high intellectual ability and operationalized as a high intelligence in practice. However, IQ is a summative measure. Although IQ gives us some indication of intellectual and academic potential, making it one of several useful gifted identification tools, it does not provide information about progress in any domain. Because of research from education and psychology on multiple domains of performance, we know much more about the nature of giftedness and how it develops. Current frameworks (e.g., Dai, 2020; Subotnik et al., 2011, 2018) emphasize several key points:

1. Giftedness is a developmental process, with individuals becoming increasingly more competent, knowledgeable, and productive within a domain as a result of support and special opportunities.
2. Different and more specific abilities are important for different domains. For example, mathematical and spatial reasoning ability are important for achievement in STEM domains. General intellectual ability—especially the capacity to learn quickly—plays a role in most domains and particularly in academic domains but becomes less relevant at higher stages of talent development.
3. Domains of talent have unique trajectories including when domain-specific abilities relevant to the domain are initially manifested and can be observed or measured. There are critical domain-specific opportunities that develop the talent and put individuals on the educational and career paths that lead to professional work in that domain. Some trajectories, such as mathematics, might begin in elementary school, and others such as diplomacy, may not be recognized or developed until early adulthood.
4. Talent or giftedness in every domain starts with potential for high achievement and, with significant and appropriate inputs within and outside of school, can be

DOI: 10.4324/9781003285991-19

developed into needed competencies, expertise, and eventually, creative productivity in adulthood.

5. Abilities are malleable and the display of ability on tests or school achievement is highly dependent on previous opportunity to learn, thus potentially disadvantaging students from low-income families. Equitable identification procedures and unique program models are needed for these students (see Olszewski-Kubilius & Corwith, 2018; Olszewski-Kubilius et al., 2020).
6. Achievement in any domain requires psychosocial skills such as motivation, the ability to embrace challenge, resilience, teachability, risk-taking, and control of performance anxiety. The transition to higher stages of talent development is often more dependent on these skills than ability (Subotnik et al., 2011).
7. Psychosocial skills are not inherent characteristics of successful individuals. They are cultivated and developed by parents, teachers, coaches, mentors, and appropriate environments (Olszewski-Kubilius et al., 2019; Subotnik et al., 2011).
8. Talent is developed via the interactions of individuals and the contexts within which they live, including family, neighborhood, community, school and society. When these contexts work synchronously to support a child's talent, it is more likely to be realized (Dai, 2020).
9. Initially, parents and teachers manage the talent development process, but with development, the individual must exert personal agency to select opportunities, activities, contexts, and individuals that support their chosen domain of talent and interest.
10. The objective of gifted programming is to move individuals along talent development paths, developing the knowledge and skills needed to progress on domain trajectories, with the goal of manifesting creative contributions in adulthood.

If our goal as a field is to enable individuals with high potential to get on talent development pathways that match their interests and abilities and stay on them, how do we know that they are making sufficient progress at each stage to successfully transition to higher stages? What are the markers or criteria that assure us that students are on track within their domain of talent? What should teachers look for in students' academic achievement and behavior to determine if they are making appropriate and significant growth in terms of their talent development? In what ways can they glean information from out of school opportunities to inform these assessments? In this chapter, we will focus on ways that teachers can assess how and whether students have acquired the knowledge and psychosocial skills needed to progress on talent development trajectories.

Talent Trajectories Defined

Giftedness or talent is not an all or nothing phenomenon nor a characteristic present within an individual at birth. Every individual comes with biological predispositions that include temperament, personality, and potential abilities (Simonton, 2005). However, the expression of these is very much dependent upon and influenced by the environment. Whether specific abilities are manifested requires opportunities to discover, display, and nurture them. Similarly, personality characteristics that support talent development such as conscientiousness or self-efficacy can be enhanced or diminished, depending upon the individuals' experiences and environment.

Csikszentmihalyi and Robinson (1986, p. 271) wrote,

> The point is that, if we agree talent depends on social attributions rather than on a naturalistic trait locked in the child's physiology, then it follows that talent should be thought of not as a stable characteristic but as a dynamic quality dependent on changes within the individual and within the environment.

Creative contributions in the form of performances, ideas, and products do not emerge in a vacuum but are the result of a multitude of influences over time as the individual grows and develops. Giftedness requires a delicate balance and synchrony of many components, including supportive factors within the family, school, community, and the broader society. Ideally, the various contexts within which students live, study, and work mutually support the development of their talent (Bronfenbrenner, 2005).

Historically, giftedness has been defined as high general reasoning ability, operationalized as high IQ. Still, today, many schools use IQ to select students for gifted services, particularly at the elementary level of schooling, often along with measures of achievement (Callahan et al., 2014). Research indicates that general cognitive ability—that is, IQ—has predictive validity for school achievement, job performance, and general life outcomes (Neisser et al., 1996; Nisbett et al., 2012; Wai et al., 2018), as well as creative productive achievements in adulthood (Lubinski, 2016).

Research also clearly tells us that there are individual differences in specific cognitive abilities that emerge by middle school (Gottfredson, 2003), and that these have implications for educational and career choices. Findings from the Study of Mathematically Precocious Youth indicate that measures of spatial, mathematical, and verbal reasoning ability obtained in middle school are highly predictive of individuals' adult choices of careers (Lubinski et al., 2006; Park et al., 2008; Wai et al., 2005). Individuals with a tilt of higher scores in math and spatial reasoning compared to reasoning tend to go into STEM fields whereas individuals with a tilt of higher scores in the verbal reasoning compared to mathematical reasoning tend to pursue the social sciences and humanities. This research demonstrates that there are specific cognitive abilities that are important for and predictive of achievement in specific domains (Bernstein et al., 2019; Makel et al., 2016; Wai et al., 2009), with implications for identification and provisions of gifted services.

The use of more specific abilities to determine potential for high achievement in a domain has long been employed in many performance areas (e.g., performing arts, sport). For example, choreographer Eliot Feld, based on years of experience building dance troupes and educating novice dance stars, identifies potential dancers around the age of eight via indicators of flexibility, body proportion, and physical memory (Subotnik, 2002). In field hockey, researchers (e.g., Elferink-Gemser et al., 2007, 2010) found that technical, tactical, and procedural skills differentiate elite from sub-elite players. A few domain-specific characteristics have also been associated with musical performance in several studies, including pitch perception (Freeman, 2000) and audiation (Ruthsatz et al., 2008), and voice teachers identified intonation, timbre, musicality, and the ability to control pitch as important factors related to singing talent (C. Watts et al., 2003).

There are some specific abilities that are important in several different fields. For example, spatial skills have been identified as important for a variety of fields including physics, chemistry, dentistry, the visual arts, and professions such as air-traffic control.

Many different types of spatial skills have been identified within this broad category, including mental rotation, visual perception, and the ability to mentally fold and unfold an object (Atit et al., 2013). Furthermore, recent research suggests that these spatial skills may be domain-specific (Uttal & Cohen, 2012), with different types of spatial reasoning required for different fields of chemistry and physics (Stieff et al., 2010), dentistry (Hegarty et al., 2009) and geoscience (Jee et al., 2009). The implications of this research are that the particular indicators of cognitive ability that educators use to identify talent for domains, including academic domains, will vary with age. General intellectual reasoning ability or measures of IQ will typically be more helpful for younger children who have had less exposure to and opportunity to manifest talent in a particular subject, and measures of more specific cognitive abilities will be more useful for determining needed services by middle school at the latest.

An exception in the academic domain can be found in mathematics. Differences in mathematical competence in the primary years may be associated with family income, teacher input, home learning, and memory capacity. However, the ability to discriminate quantities, discern number patterns, and rule out unreasonable results—labeled *number sense*—is an unlearned competence, and the link between number sense and math ability is present before the beginning of formal math instruction (Libertus et al., 2011). A longitudinal, multisite study with 4.5-year-old children conducted by T. W. Watts et al. (2014) showed that even after accounting for family characteristics and cognitive skills, differences on mathematics subtests of the Woodcock Johnson III Tests of Achievement could be found among participating children that predicted mathematics achievement through age 15. The children who showed exceptional growth in mathematical ability between 4.5 years and their entry into first grade were most likely to demonstrate high achievement in high school mathematics. Also, using subtests of the Woodcock Johnson battery, Siegler et al. (2012) showed that even after controlling for other types of mathematical knowledge, general intellectual ability, working memory, and family income and education, young students' mastery of fractions and division predicts both algebra and overall mathematics achievement in high school five or six years later.

Clearly, the process of developing talent also varies by domains. Whether a trajectory begins in early childhood versus adolescence, for example, depends on when the skills and abilities in the talent area emerge and coalesce enough to allow for recognition in some reliable fashion. The trajectory is affected by physical maturation in domains such as music and sports, and it also depends on when talent can be ascertained by systematic identification procedures (e.g., testing or ratings by knowledgeable adults such as educators and coaches). As indicated above, precocity in mathematical reasoning may be obvious as early as the preschool years and some children read fluently well before school entry. Interest and talent in science is often observable in middle school, whereas aptitude for the social sciences or humanities may not be recognized until late high school or college.

End points of developmental trajectories also vary widely. Trajectories can be short in most sport fields, but for most academic fields and some musical fields, these developmental arcs are virtually lifelong. In academic areas, individuals can remain involved and active well into late adulthood, with almost no limits on productivity. Intervals between starts and peaks also vary greatly by domain, with most academic fields requiring long periods of preparation (Simonton, 2007), whereas in some others (e.g., mathematics), major contributions often occur much earlier, whether by virtue of early identification, or because extended years of training are not required, such as in software engineering (McWilliams et al., 2019). These varied arcs for different domains have implications for when identification and intervention can occur and fortunately, for most academic domains, there is ample opportunity for "late bloomers," suggesting that educators need to be open to continuously discovering and nurturing talent in academic domains.

The developmental course of domain trajectories is also affected by training, educational requirements, and tradition, which are tied to our schooling system in many academic areas. For example, the serious study of social sciences, such as sociology, anthropology, or psychology, typically does not start until high school or college, although it could begin earlier, especially for verbally gifted students. As a result, specialization can typically get underway only in college, although educators can provide early opportunities in these subjects for gifted students. Domains such as psychology, religion, diplomacy, or literature often require the accumulation of maturity and experience to generate important contributions and so, typically occur later.

Although the paths that individuals take toward their adult careers is highly varied, there is some research to suggest that particular experiences and opportunities (or lack thereof) can be critically important to whether talent is developed or not. Wai et al. (2005) found that a higher dose of STEM activities in the middle and high school years—including school courses and outside of school opportunities such as summer programs, extra-curricular clubs and competitions—was related to higher levels of achievement within STEM fields. Subotnik et al. (1993) found that winners of a prestigious science fair competition were more likely to stay in university STEM majors if their high school competition projects were conducted as part of an apprenticeship in a research laboratory rather than on their own or at school. Additionally, opportunities to do authentic research in a university research lab while in high school was particularly important to gifted females' commitment to pursuing STEM careers (Subotnik et al., 2013).

Research also supports positive impacts of outside of school programs for gifted students (see Olszewski-Kubilius & Lee, 2004, 2008 for reviews), including effects that extend to their school achievement, such as taking a more rigorous course of study, greater use of accelerative options in mathematics, greater participation in math related extra-curricular activities, higher educational aspirations, and matriculation at more academically selective institutions of higher education (Barnett & Durden, 1993; Olszewski-Kubilius, 2015; Olszewski-Kubilius & Grant, 1996; Olszewski-Kubilius & Lee, 2004, 2008; Swiatek & Benbow, 1991). Understanding the nature of talent development trajectories in different domains is critical so that important and optimal opportunities and support can be provided at critical times in an individual's pathway.

Assessments for Potential

All talents begin with potential for later achievement. In some domains, that potential can manifest early—such as in the domain of mathematics—with young children exhibiting early interest in "mathematizing" their world (Krutetskii, 1976). This early precocity may be fueled by parental input or inherent proclivities and these children enter school with advanced mathematical knowledge and reasoning ability that requires early acceleration. This precocity may be obvious to teachers but may still require formal assessment via achievement tests to determine how far above grade-level the child should be placed in terms of mathematics instruction. Any standardized, well-documented achievement test that includes subtests in mathematics can be used. Tests such as the Iowa Test of Basic Skills can be used but may need to be given above grade level to ensure there is enough ceiling to measure reasoning abilities that are significantly advanced (see Table 14.1).

TABLE 14.1 Assessments for different talent development levels

Potential	
Demonstration of above age typical ability-precocious reading, mathematizing external world.	Formal assessments in some areas such as reading or math; informal in others such as science.
Enthusiasm for learning activities.	Observed by parents and teachers.
Pushing adults to provide learning activities.	Reported by parents.
Demonstration of high interest—questions about natural world.	Observed by parents and teachers.
Ability to sustain attention beyond age-appropriate expectations.	Observed by parents and teachers.
Scores on individual or group tests of ability	Assessments by teachers or school psychologists.
Competency	
Demonstration of above average reasoning ability in domain.	Formal performance via tests, informal assessment via class performance.
Ability to sustain attention with challenging problems, assignments.	Observations of teachers; quality of class assignments.
Openness to feedback and critique.	Observations of teachers; evidence of revision of work based on feedback.
Developing organizational skills.	Assignments handed in on time; evidence of planning for longer-term projects, quality of projects and assignments.
Evidence of use of learning strategies for challenging work.	Self-report of study skills for different areas.
Burgeoning knowledge of strengths and weaknesses.	Self-reflection on areas to improve upon and work on.
Evidence of agency in finding materials of interest in the domain	Self-report.
Openness to challenge, competition.	Teacher observation of attention to and enthusiasm for challenge,
Ability to work well with others.	Teacher observation.
Expertise	
Demonstration of superior knowledge, performance within the domain.	Formal assessments via off-level tests in domain; grades.
Demonstration of interest and agency in pursuit of learning activities related to the domain.	Patterns of participation in course selection extra-curricular, and outside of school activities, independent, self-initiated projects.
Realistic appraisal of strengths and weaknesses; agency in shoring up weaknesses.	Evidence of pursuit of activities to develop needed ancillary skills—e.g., speaking, writing, presentation skills.
Exploration of domain-related career and educational paths.	Evidence of consultation with school counselors, teachers.
Willingness to engage in competitive activities.	Participation in contests, submission of work for publication, etc.
Evidence of growing domain-specific scholar identity.	Patterns of participation in course selection extra-curricular, and outside of school activities, independent, self-initiated projects.

Computer-based, adaptive tests, such as the Measure of Academic Progress or MAP test (www.nwea.org) can also be used to measure reasoning above grade level. A third alternative is an individually administered, standardized achievement test (e.g., the Wechsler Individual Achievement Test or the Woodcock-Johnson Tests of Achievement), as these measures have items that assess academic knowledge across the full K–12 range and can provide a sufficient ceiling for assessing advanced young learners.

Competency-based tests should not be used for this purpose as they will only provide information regarding students' knowledge of grade-level mathematics, which is not useful for appropriate placement. Additionally, if the school uses a mathematics curriculum that has an end-of-year assessment, this assessment can be used to identify skill gaps, if any, that might need to be addressed prior to an accelerated placement. Skill gaps on specific topics should not prohibit an accelerated placement and should be expected. The Iowa Acceleration Scale, 3rd Edition (Assouline et al., 2009), which provides a way of compiling information on a child's ability and readiness for acceleration, can be used to make decisions about subject area or whole-grade acceleration.

Other children may also have potential talent in mathematics that is not as obvious due to less exposure or opportunity. At the potential stage, these children will benefit from enrichment opportunities with built-in assessments involving challenging math problems, so that teachers can notice their exceptional mathematical reasoning ability and nurture it with differentiated instruction. A good curriculum for high level, conceptually oriented enrichment is Mentoring Young Mathematicians (https://gifted.uconn.edu/projectm2/). An example of this approach in practice is the Young Scholars Program (Horn, 2015) in which teachers are trained to offer challenging lessons to early elementary students and notice exceptional reasoning ability. Children are then monitored and provided continuing opportunities to develop their abilities, with the goal of enabling them to enter school and district-based gifted programs that require specific performance criteria. For these children, services in the form of challenging lessons come first, followed by later assessment via standardized achievement tests (Olszewski-Kubilius et al., 2018).

At the potential stage, interest and engagement can be a key to identifying advanced potential if children are given opportunities to develop and showcase their talents. Interests can also be determined by careful observation of the activities that children engage in when given a choice, the questions they ask, and the persistence demonstrated during domain specific lessons. Teachers need to be making and documenting observations over time. Additionally, an observation scale, such as the Scales for Identifying Gifted Students (SIGS-2), can be used by teachers to rate students in specific content areas—once opportunities have been provided. Additionally, although most of the early elementary curriculum is focused on literacy and mathematics, lessons in science should be provided with similar attention to student interest and exceptional reasoning ability.

Similarly, some students will enter school already reading above grade level. Formal assessment via standardized tests that assess decoding and reading comprehension as well as other language skills will be needed to determine what adjustments to the curriculum should be made, including accelerated placements. Exceptional verbal reasoning ability may also manifest in reading nonfiction books on science or other topics, which can be observed via students' reading choices. Most formal reading tests assess decoding, vocabulary, and reading comprehension, but teachers should supplement with informal assessments of other aspects of verbal reasoning ability such as writing tasks, via in-classroom, differentiated enrichment activities provided by the teacher. These supplemental tasks can be used to determine skill levels and adjust instruction to promote further development. Although early reading fluency is not a definitive predictor of gifted performance, it does indicate the need for more advanced learning opportunities that require reading skills.

Although some children will demonstrate potential in specific subject areas, other students will demonstrate advanced general reasoning ability, but not necessarily advanced knowledge in a particular subject. Informally, this latter group of students may be observed by teachers to be "fast learners," requiring lessons differentiated for pacing and depth. If possible, measures of general reasoning ability can be used to universally screen all students for advanced learning potential. The Cognitive Abilities Test (CogAT; www.riversideinsights.com) yields measures of verbal, nonverbal, and spatial reasoning ability. Alternatively, some students may require individualized cognitive tests to showcase their potential. It is important to monitor students who demonstrate high general reasoning ability for continued engagement and enthusiasm for learning, as motivation can be lost due to inappropriate placement, a learning pace that is too slow, or learning activities focused on basic skills. Keeping in mind the varied trajectories of domain talent, many students at this stage will not exhibit a preference for or advanced ability in one domain over another, and talent may remain generalized until middle or high school. Continuous monitoring is particularly important for high achievers from lower income families, who, based on research, are more likely to fall out of high achievement levels as they proceed through school (Wyner et al., 2007).

In summary, at the potential stage, both formal assessment for determination of placement in mathematics and reading and informal assessment via the provision of challenging learning activities are essential. In addition, at this stage, parental input is important and useful (the SIGS-2 has a home scale for parents to rate their children). Teachers should seek information from families on what their students choose to read at home, what they choose to pursue independently at home (e.g., writing plays, doing science experiments, working math problems, drawing) or via computer programs, websites, and other media, and what family activities they exhibit great interest and enthusiasm for (e.g., trips to the science museum). Research shows that some students who are disengaged in school often spend considerable time within online communities to learn how to make movies, get feedback on their writing, or do service projects (Ito et al., 2013), and one way to reengage them is to bring those interests more deliberately into the classroom. At this initial stage of talent development, parents and teachers need to work together to insure they have sufficient information on their students regarding their talent areas and interests so that they can be acknowledged and nurtured within school.

Assessments for Psychosocial Skills

Additionally, as noted above, psychosocial skills are important to the fruition of talent. These skills include learning to persist, stay motivated, and channel competition into self-improvement. Psychosocial skills are also developed through appropriate opportunities and should not be used to determine students' eligibility for gifted services. Teachers should note students' strengths and areas for improvement, however, and adjust messages and feedback to help cultivate these skills. Teachers can promote growth mindsets by the feedback they give to students that emphasize growth and learning goals rather than performance goals (Ricci, 2017; Sanguras, 2017). They can help students acquire resiliency and persistence by encouraging and supporting them as they work through challenging lessons or difficult problems (Mofield, 2018). At the stage of potential, there should be considerable latitude for the presence or absence of these psychosocial skills, with a greater emphasis on developing them rather than evaluating them.

Assessments for Competence

Building foundational knowledge and skills is the goal at the competency stage of talent development. It is important that growth is carefully monitored through formal assessments, particularly by using above-grade-level achievement tests, as many children identified as gifted will already demonstrate high levels of performance (Peters et al., 2017). Grade-level tests can demonstrate mastery of grade-level curriculum but will not be able to show growth or mastery of above-grade-level material due to ceiling effects. There is ample research to show the efficacy of using the ACT, PSAT, and SAT with middle school students to assess level of knowledge and skills in the verbal and mathematical domains (Olszewski-Kubilius, 2015). These tests have an adequate ceiling to be helpful in determining subject area placement and the ACT has the advantage of also including a science reasoning subtest. They can be used to assess growth, which should be accelerated if students are being provided appropriate levels of curriculum and instruction. Although some students may show above-grade-level performance across multiple subjects, teachers should expect to see differential performance and growth for areas of talent compared to relatively weaker areas. Monitoring growth, rather than absolute performance, is particularly important so that students do not lose ground or disengage.

Informal assessments of other skills are important at this stage of talent development as well. A critical psychosocial skill at the competency stage is openness to feedback and critique by teachers, as this is how students improve and grow and how teachers reinforce growth mindsets. Teachers can observe whether students respond to feedback regarding suggested improvements on projects and assignments, which can be assessed via students' portfolios. Do revisions of a writing assignment show responsiveness to feedback and improvement? Do power point presentations improve as a result of suggestions? Do students' projects show increased depth as a result of teacher recommendations? A good resource for determining growth (via rubrics) on many different types of projects is the second edition of a book by Roberts and Inman (2015): *Assessing Differentiated Student Products: A Protocol for Development and Evaluation*. Of course, being able to assess growth on students' skills is dependent upon teachers periodically giving advice and feedback for improvement as students work on longer term, larger projects.

Teachers can observe whether students are willing and motivated to take on more challenging work. This motivation may be evident in persistence when presented with more difficult assignments or problems, or in the selections students make when they are able to choose topics to study or the way in which they will demonstrate their learning. Are they choosing more complex topics to study consistent with their abilities or selecting ways to demonstrate learning that stretches their skills (e.g., creating a podcast or movie versus a power point)?

Teachers can help students reflect on their study habits and use of learning strategies, particularly when given more challenging work. Have students increased their effort in more challenging classes? Are students acquiring and applying new learning strategies for more difficult material, and are these changes resulting in improved performance and learning? Do students' portfolios show increased adeptness at using organizational skills, such as breaking large assignments and projects into steps that are completed well and on time, or are students waiting until the last minute to complete assignments and projects with substandard results? Do students use planners or other organizational tools regularly to track assignments and larger projects? Additionally, are students showing evidence that they are acquiring independent learning skills, such as pursuing information on a topic of interest or a topic needed for an assignment on their own? Other indications of burgeoning autonomous learning skills can include information about projects pursued at home or outside of school.

At the competency stage, students can learn to recognize and itemize their strengths and weaknesses to capitalize on talents but also work to shore up weaker areas. However, students at this stage are still developing, and perceptions might be influenced heavily by the kinds of opportunities they have had. Care should be taken to not overly "pigeonhole" students, while assisting them in acknowledging their talents, recognizing their interests, and working to develop other areas that will support continued talent development. Both scientists and artists also need good speaking and writing skills. Other "uber" skills that enhance success in any specific talent area include research skills and constructive responsiveness. Teachers can also observe whether students are able to work successfully with others. Via informal observation, they can determine whether students can assume different roles in different groups, work cooperatively on group projects and assignments, interact appropriately with others when sparring over or sharing ideas, and contribute productively without dominating. School psychologists can be called upon to help students learn how to assume varied roles for group projects, temper a tendency to dominate, or develop confidence to speak up and share ideas.

Unfortunately, many of the psychosocial skills do not have formal assessments, and their assessment will rely on teacher report and observations. School psychologists and social workers can provide assessments for concerns about learning disabilities, impairments in executive function skills, or social skill impairments that significantly affect performance. Teachers can also use an affective curriculum, such as *Teaching Tenacity, Resilience, and a Drive for Excellence: Lessons for Social-Emotional Learning for Grades 4-8* (Mofield, 2018), which provides activities to reinforce positive psychosocial skills for high achievement, but can also serve as a way for teachers to assess them and identify students who may need some assistance with developing better skills.

Assessments for Expertise

The expertise stage typically begins as early as the later years of secondary school and continues through graduate training and beyond; in some fields, training continues after formal classroom-based education is completed (e.g., the internship, residency, and fellowship in medicine). Above grade level testing may still help drive decisions at the end of secondary school but even typical college entrance exams will likely exhibit ceiling effects for many students. For example, many gifted students at this stage have taken multiple college courses. At this stage of development, what students actually *do* with their ability and around their interests is more important than test scores. Teachers will benefit from the advice of adult domain experts to determine if students are on track to meet domain standards of preparedness and achievement. For example, medical doctors can help teachers assess whether students are gaining the educational and experiential portfolio needed to gain entry into medical school and engineers can advise whether students have the necessary courses required for entrance into selective university engineering programs. Writers and journalists can help advise students on how to prepare for selective university journalism programs. It is never too early to get advice on educational and career paths, but this should happen no later than high school.

Counselors and teachers should review students' academic profiles to see if there are patterns of increasing competence in the area of talent. Has the student taken a substantial portion of the courses available from their school within their talent area? Are they taking advantage of the advanced programs that the school offers such as IB and dual credit programs and AP classes? Are they succeeding in these more challenging courses and programs in terms of grades or performance on AP tests? Is their performance in other areas at a sufficiently high level to enable them to qualify for selective college programs in their talent area?

At this stage of talent development, demonstration of increasing agency on the part of individuals in directing and managing their own talent development is important. This increase in autonomy can be assessed informally by teachers and counselors via conferencing with students. Is the student selecting courses that match his or her interests and talent area? Is there evidence of independent pursuit of interests such as participation in extra-curricular activities including school clubs, service organizations, mentorships, internships, apprenticeships, and competitions, and are these complementary to strengths or addressing relatively weaker areas (e.g., debate to improve speaking)? Are students seeking available outside-of-school opportunities or asking for assistance from school personnel to access these in their talent area? Additionally, are students demonstrating active exploration of educational paths towards careers they are interested in or asking for assistance from school advisors and counselors? Not only can this information inform teachers and counselors if students are on talent development trajectories, but also this information can reveal gaps in students' capacity to seek out support and assistance. Knowledge about how, where, and when to ask for help is especially important for students who are first generation college attendees, and they may need assistance accessing this information from counselors or professionals in the community.

One of the ways in which counselors and teachers can assist students is to help them build a personal profile or portfolio for tracking their coursework, special project work, independent activities outside of school in their talent area, extra-curricular activities, and any outside-of-school programming they participated in. These portfolios help to provide a picture of a student's growing expertise and commitment and should be reviewed periodically by students and teachers to identify missing components or under-developed areas that need attention (see Kay, 2019).

Teachers and counselors will need to be especially attentive to students who have been under-represented in advanced courses to ensure they are being counseled into these programs and being successful in them. Research suggests that culturally and linguistically diverse and low income students are at risk for passing up on advanced courses, even when they are available within their schools and they have the skills to be successful (Theokas & Bromber, 2014; Wyner et al., 2007). Monitoring, support, and collaboration with teachers, advisors, and college counselors are needed. Undermatching college choices is common among lower-income gifted students, who fear the high costs of private institutions and lack knowledge about the potential for substantial financial assistance (Theokas & Bromber, 2014). These students and their families require early and effective college counseling to ensure students choose supportive, appropriately challenging institutions of higher education.

Conclusion

Talent development is a long-term process. As such, supports should be ongoing and articulated across levels of schooling. As articulated above, assessment is much more than an end of semester test. Rather, assessment refers to "a process that integrates test information with information from other sources," such as observations, educational histories (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Educators must take the perspective that they are accountable for ensuring that students acquire the knowledge and have the critical domain-relevant educational experiences that enable them to transition to the next stage of talent development. Additionally, equal attention must be given to helping students acquire supportive psychosocial skills as content knowledge, as lack of these skills is what typically derails students at higher stages.

Resources

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Assouline, S., Colangelo, N, Lupkowski-Shoplik, A., Lipscomb, J., & Forstadt, L. (2009). *Iowa Acceleration Scale* (3rd ed.). Great Potential Press.

Johnsen, S. K., & Corn, A. L. (2018). Screening Assessment for Gifted Elementary and Middle School Students (SAGES-3). PRO-ED.

Klinger, D. A., McDivitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., & Wylie, E. C. (2015). *The classroom assessment standards for PreK–12 teachers*. Kindle Direct Press.

Mofield, E. (2018). *Teaching tenacity, resilience, and a drive for excellence; Lessons for social-emotional learning for Grade 4–8*. Prufrock Press.

Neihart, M. (2008). *Peak performance for smart kids: Strategies and tips for ensuring school success*. Prufrock Press.

Ricci, M. C. (2017). *Mindsets in the classroom; Building a growth mindset learning community*. Prufrock Press.

Roberts, J. L., & Inman, T. F. (2015). *Assessing differentiated student products: A protocol for development and evaluation* (2nd ed). Prufrock Press.

Sanguras, L. (2017). *Grit in the classroom. Building perseverance for excellence in today's classrooms*. Prufrock Press.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Assouline, S., Colangelo, N., Lupkowski-Shoplik, A., Lipscomb, J., & Forstadt, L. (2009). *Iowa Acceleration Scale* (3rd ed.). Great Potential Press.

Atit, K., Shipley, T. F., & Tikoff, B. (2013). Twisting space: Are rigid and non-rigid mental transformations separate spatial skills? *Cognitive Processing*, *14*(2), 163–173. https://doi.org/10.1007/s10339-013-0550-8

Barnett, L. B., & Durden, W. G. (1993). Education patterns of academically talented youth. *Gifted Child Quarterly*, *37*(4), 161–168. https://doi.org/10.1177/001698629303700405

Bernstein, B. O., Lubinski, D., & Benbow, C. P. (2019). Psychological constellations assessed at age 13 predict distinct forms of eminence 35 years later. *Psychological Science*, *30*(3), 444–454. https://doi.org/10.1177/0956797618822524

Bronfenbrenner, U. (2005). *Ecological systems theory (1992).* In U. Bronfenbrenner (Ed.), *Making human beings human: Bioecological perspectives on human development* (pp. 106–173). Sage Publications.

Callahan, C. M., Moon, T. R., & Oh, S. (2014). *National surveys of gifted programs: Executive summary 2014*. Curry School of Education, University of Virginia. www.nagc.org/sites/default/files/key%20reports/2014%20Survey%20of%20GT%20programs%20Exec%20Summ.pdf

Csikszentmihalyi, M., & Robinson, R. E. (1986). Culture, time and the development of talent. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 264–284). Cambridge University Press.

Dai, D. Y (2020). Evolving complexity theory of talent development: A developmental systems approach. In T. L. Cross & P. Olszewski-Kubilius (Eds.), *Conceptual frameworks for giftedness and talent development. Enduring theories and comprehensive models in gifted education* (pp. 1–28). Prufrock Press.

Elferink-Gemser, M. T., Kannekens, R., Lyons, J., Tromp, Y., & Visscher, C. (2010). Knowing what to do and doing it: Differences in self-assessed tactical skills of regional, sub-elite, and elite youth field hockey players. *Journal of Sports Sciences*, *28*(5), 521–528. https://doi.org/10.1080/02640410903582743

Elferink-Gemser, M. T., Visscher, C., Lemmink, K. A. P. M., & Mulder, T. (2007). Multidimensional performance characteristics and standard of performance in talented youth field hockey players: A longitudinal study. *Journal of Sports Sciences*, *25*(4), 481–489. https://doi.org/10.1080/02640410600719945

Freeman, J. (2000). Children's talent in fine art and music. *Roeper Review, 22*(2), 98–101. https://doi.org/10.1080/02783190009554010

Gottfredson, L. S. (2003). The science and politics of intelligence in gifted education. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 24–40). Pearson.

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*(1), 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Horn, C. V. (2015). Young scholars: A talent development model for finding and nurturing potential in under-served populations. *Gifted Child Today*, *38*(1) 19–31. https://doi.org/10.1177/1076217514556532

Ito, M., Gutiérrez, K. D., Livingstone, S., Penuel, B., Rhodes, J. E., Salen, K., Schor, J. Sefton-Green, J., & Craig Watkins, S. (2013). *Connected learning: An agenda for research and design*. Digital Media and Learning Research Hub. https://clalliance.org/publications/connected-learning-an-agenda-for-research-and-design/

Jee, B., Gentner, D., Forbus, K., Sageman, B., & Uttal, D. H. (2009). Drawing on experience: Use of sketching to evaluate knowledge of spatial scientific concepts. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam, The Netherlands*.

Kay, S. (2019). *On human potential: Nurturing talents and cultivating expertise*. Rowman & Littlefield.

Krutetskii, V. A. (1976). *The psychology of mathematical abilities in school children*. University of Chicago Press.

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

Lubinski, D. (2016). From Terman to today: A century of findings on intellectual precocity. *Review of Educational Research*, *86*(4), 900–944. https://doi.org/10.3102/0034654316675476

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science 17*(3), 194–199. https://doi.org/10.1111/j.1467-9280.2006.01685.x

Makel, M., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science*, *27*(7), 1004–1018. https://doi.org/10.1177/0956797616644735

McWilliams, M. A., Holding, E. Z., & Knotek, S. E. (2019). Talent development in medicine and software engineering. In R. F. Subotnik, P. Olszewski-Kubilius, & F. C. Worrell (Eds.), *The psychology of high performance: Translating human potential into domain-specific talent* (pp. 109–134). American Psychological Association. https://doi.org/10.1037/0000120-006

Mofield, E. (2018). *Teaching tenacity, resilience, and a drive for excellence; Lessons for social-emotional learning for Grade 4–8*. Prufrock Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Nisbett, R. E., Aronson, J., Blair, C. Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Olszewski-Kubilius, P. (2015). Talent searches and accelerated programming for gifted students. In S. Assouline, N. Colangelo, J. VanTassel-Baska, & A. Lupkowski-Shoplik (Eds.), *A nation empowered: How schools hold back America's brightest students* (pp. 111–122). University of Iowa.

Olszewski-Kubilius, P. & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly*, *62*(1), 37–55. https://doi.org/10.1177/0016986217738015

Olszewski-Kubilius, P., & Grant, B. (1996). Academically talented females and mathematics: The role of special programs and support from others in acceleration, achievement and aspirations. In K. Arnold, K. Noble, & R. F. Subotnik (Eds.), *Remarkable women: Perspectives on female talent development* (pp. 281–294). Hampton Press.

Olszewski-Kubilius, P., & Lee, S.Y. (2004). Gifted adolescents' talent development through distance learning. *Journal for the Education of the Gifted*, *28*(1), 7–35. https://doi.org/10.1177/016235320402800102

Olszewski-Kubilius, P., & Lee, S. Y. (2008). Specialized programs serving the gifted. In F. A. Karnes & K. P. Stephens (Eds.), *Achieving excellence. Educating the gifted and talented* (pp. 192–208). Pearson Education.

Olszewski-Kubilius, P., Stambaugh, T., & Corwith, S. (2020). Poverty, academic achievement and talent development. In T. Stambaugh & P. Olszewski-Kubilius (Eds.), *Unlocking potential. Identifying and serving gifted students from low-income households* (pp. 3–26). Prufrock Press.

Olszewski-Kubilius, P. Subotnik, R. F., Davis, L. C., & Worrell, F. C. (2019). Benchmarking psychosocial skills important for talent development. *New Directions for Child and Adolescent Development*, *168*, 161–176. https://doi.org/10.1002cad.20318

Olszewski-Kubilius, P., Subotnik, R. F., & Worrell, F. D. (Eds.). (2018). *Talent development as a framework for gifted education: Implications for best practices and applications in schools*. Prufrock Press.

Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, *19*(10), 957–961. https://doi.org/10.1111/j.1467-9280.2008.02182.x

Peters, S. J., Rambo-Hernandez, K., Make., M. S., Matthews, M. S., & Plucker, J. A. (2017). Should millions of students take a gap year? Large numbers of students start the school year above grade level. *Gifted Child Quarterly, 61*(3), 229–238. https://doi.ore/10.1177/0016986217701834

Ricci, M. C. (2017). *Mindsets in the classroom; Building a growth mindset learning community*. Prufrock Press.

Roberts, J. L., & Inman, T. F. (2015). *Assessing differentiated student products: A protocol for development and evaluation* (2nd ed). Prufrock Press.

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence*, *36*(4), 330–338. https://doi.org/10.1016/j.intell.2007.08.003

Sanguras, L. (2017). *Grit in the classroom. Building perseverance for excellence in today's classrooms.* Prufrock Press.

Siegler, R. S, Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Ines Susperreguy, M., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*(7), 691–697. https://doi.org/10.1177/0956797612440101

Simonton, D. K. (2005). Giftedness and genetics: The emergenic-epigenetic model and its implications. *Journal for the Education of the Gifted*, *28*(3–4), 270–286. https://doi.org/10.4219/jeg-2005-338

Simonton, D. K. (2007). Creative life cycles in literature: Poets versus novelists or conceptualists versus experimentalists? *Psychology of Aesthetics, Creativity and the Arts*, *1*(3), 133–139. https://doi.org/10.1037/1931-3896.1.3.133

Stieff, M., Hegarty, M., & Dixon, B. (2010). Alternative strategies for spatial reasoning with diagrams. *Diagrams*, *2010*, 115–127. https://doi.org/10.1007/978-3-642-14600-8_13

Subotnik, R. F. (2002). Talent developed—Conversations with masters in the Arts and Sciences. Eliot Feld: Innovator in choreography and dance company development. *Journal for the Education of the Gifted*, *25*(3), 290–302. https://doi.org/10.1177/016235320202500305

Subotnik, R. F., Duschl, R. A., & Selmon, E. H. (1993). Retention and attrition of science talent: A longitudinal study of Westinghouse Science Talent Search winners. *International Journal of Science Education*, *15*(1), 61–72. https://doi.org/10.1080/0950069930150105

Subotnik, R. F., Olszewski-Kubilius, P. & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2018). Talent development as the most promising focus of giftedness and gifted education. In S. Pfeiffer, E. Shaunessy-Dedrick, & M. Foley-Nicpon (Eds.), *APA handbook of giftedness and talent* (pp. 231–245). American Psychological Association. https://doi.org/10.1037/0000038-015

Subotnik, R. F., Tai, R. H., Almarode, J., & Crowe, E. (2013). What are the value added contributions of selective secondary schools of mathematics, science, and technology? Preliminary analyses from a U.S. national research study. *Talent Development and Excellence*, *5*, 87–97.

Swiatek, M. A., & Benbow, C. P. (1991). Ten-year longitudinal follow-up of ability matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology*, *3*(4), 528–538. https://doi.org/10.1037/0022-0663.83.4.528

Theokas, C., & Bromber, M. (2014). *Falling out of the lead: Following high achievers through high school and beyond*. The Education Trust. https://edtrust.org/resource/falling-out-of-the-lead-following-high-achievers-through-high-school-and-beyond/

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why and how? *Psychology of Learning and Motivation*, *57*, 147–181.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, *97*(3), 484–492. https://doi.org/10.1037/0022-0663.97.3.484

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Wai, J., Worrell, F. C., & Chabris, C. (2018). The consistent influence of general cognitive ability in college, career, and lifetime achievement. In K. Larsen-McClarty, K. Mattern, & M. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice* (pp. 46–56). Routledge. https://doi.org/10.4324/9781315621975-5

Watts, C., Barnes-Burroughs, K., Andrianopoulos, M., & Carr, M. (2003). Potential factors related to untrained singing talent: A survey of singing pedagogues. *Journal of Voice*, *17*(3), 298–307. https://doi.org/10.1067/S0892-1997(03)00068-7

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352–360. https://doi.org/10.3102/0013189X14553660

Wyner, J. S., Bridgeland, J. M., & DiIulio Jr, J. J. (2007). Achievement trap: How America is failing millions of high-achieving students from lower-income families. *Civic Enterprises*.

Conclusion to Part II on Assessments for Learning Progress

Joyce VanTassel-Baska and Susan K. Johnsen

Introduction

The chapters contained in Part II constitute important perspectives on the use of learning assessments with gifted learners. What makes an assessment appropriate to judge the outcomes of advanced learning for different program prototypes? Why does no one assessment provide all of the criteria for judging student outcomes? If classroom tools are effective in making such judgments, why are they not used more consistently? The chapters represent points of view on program interventions as well. The program model, the outcomes and expectations identified for learners, and the potential uses of the assessments all impact the tools selected for the task. Sometimes the tools themselves are fraught with technical adequacy problems when teachers "create" assessments without validity and reliability verification, processes that require collaboration with other teachers and assistance from staff. Hopefully, the lessons learned from these chapters will constitute a guide for gifted educators in their quest to find best practices for assessing advanced learning at all stages of development.

Figure II.1 shows the cyclical process of developing learning assessments for advanced learners. Assessment tools must be selected that meet technical adequacy standards and that may be administered pre-post in order to identify strengths and needs and to document learning changes. If the tool selected is post-only, then other pre-assessment data should be collected to understand the nature of learning progress and future needs. Learning outcomes must be established that are advanced yet reflect an alignment to the relevant curriculum standards. Once this framework is established, the details of the intervention might be planned to reflect activities, materials, and projects selected to address the learning outcomes. The teacher then would facilitate the learning process to enable growth to occur in the advanced content, higher level processes, and authentic products that learners might experience and create. After sufficient time for the intervention (i.e., a unit of study or a course), the post-assessment would be administered and scored, using a rubric that has been judged technically sound or paper-pencil posttest for a more standardized measure. Results from the

DOI: 10.4324/9781003285991-20

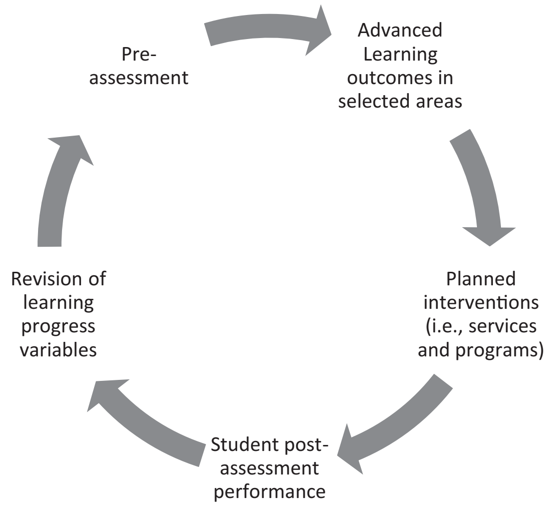

FIGURE II.1 The learning progress cycle in gifted programs and services

assessment phase of the process would be analyzed for learner growth overall and sub-analyzed by underrepresented groups. Based on these results and input from stakeholders (i.e., students and teachers), revisions may be made to relevant aspects of the cycle. For example, the learning outcomes may not be advanced enough or not well-aligned to the standards to be addressed or the assessment tools may be too rigorous for the intervention offered or the time expended. Moreover, the learning activities might not be aligned to student interests, suppressing student motivation and performance. Consequently, specific revisions may need to be made within the implementation of the cycle to align to curriculum standards more closely and to each student's strengths, needs, and interests.

Summaries of Chapters

The following encapsulation of these seven chapters offers a window into different types of learning assessments, beginning with those that are traditional and standardized and proceeding to those that offer alternative approaches that vary from performance-based to self-assessment. Each provided a rationale and research support for using the approach discussed and offered examples of what it looks like and how it might be implemented in schools or classrooms. Each author or set of authors has both university expertise and school experience in developing and implementing learning assessments for gifted learners. Thus, the views presented are grounded in both research expertise and classroom experience.

The Boswell et al. chapter analyzed the value of primarily *state assessments* in gauging the advanced learning of gifted students. The authors began by presenting the problems of using these assessment tools in isolation, especially due to the problem of low ceilings and lack of consistency across states. The authors then analyzed state data available through the State of the States Report of NAGC for 2020 to examine the processes being employed to assess gifted learners appropriately on state tests. Only 11 states reported the status of differentiated assessment for the gifted as a subgroup, with each of the 11 using different approaches to the concept of "growth assessment" of gifted learners. Within that small number of states, there were some notable examples. Ohio disaggregates achievement scores for different gifted subgroups, examining how gifted students are identified and served. All of these data points are quantified to produce a Gifted Indicator score that reflects how well a school or district is meeting the expectations for gifted performance and progress. Colorado offers growth assessment through Advanced Learning Plans. While these plans rely on test scores, they are a unique approach to identifying learning needs of gifted students at the local level. Texas reports the use of an open-ended, project-based option for assessing K- 12 gifted learners' progress (i.e., Texas Performance Standards Project). The project's rubric covers two general categories: developing and implementing a research plan (the research process) and presentation of learning (product and communication). The authors concluded their chapter with a set of recommendations for states in respect to a continued lack of attention to policy and its enactment: a need to disaggregate gifted learners as a subgroup in order to analyze annually their growth in learning, the use of a growth model for the gifted subgroup that examines that learning, and the need for using alternative assessments in addition to existing state ones.

The Brown chapter focused on the use of assessments in secondary programs for gifted students. She cautioned educators about using the traditional state tests alone to judge advanced levels of learning and about the need to use a mix of types of assessment to make such judgments. This chapter provided concrete descriptions and specific examples of Advanced Placement and International Baccalaureate programs and assessments that are widely used across the country to provide advanced credit, placement, and standing to students who access them. Other *secondary performance-based assessments*, used to assess learning within curriculum units of study or across a year of coursework in selected content areas, were also highlighted with relevant examples provided. Problem-based learning assessments were also noted as another tool to apply to judging advanced thinking and problem-solving. The role of competitions was also described as a way of documenting growth in an area of learning. The chapter concluded with an emphasis on administrator strategies needed to be considered in using these alternative forms of assessment in gifted programs at all levels of the secondary experience.

The Kettler and Lamb chapter engaged with the topic of creativity as the overall goal of gifted education and how to assess it in the context of school learning that is subject-specific and driven by standards. The authors debunked the idea that creativity is too complex to be assessed, noting that it is content-specific and should be assessed accordingly. The authors explicated how *creativity assessment* may occur at the individual level (person), the process level, and the product level. They highlighted the evidence for the power of assessing the processes that enhance creative work. Meta-analytic research, conducted on 70 empirical studies on the effectiveness of creativity training, has found that creative processes can be improved in students and adults through specific training and practice. This finding buttressed the authors' contention that all core content might be infused with creative skills and creative problem-solving tactics to enhance the learning of the gifted. Finally, the authors explored the use of product assessment through the Consensual Assessment Techniques (CAT), applied to judge products in different domains and using expert reviewers from those domains as the judges. The authors concluded their chapter by reiterating that the development of creative thinking and problem-solving skills is fundamental to talent development and exceptional academic achievement.

These next three chapters were organized to address learning assessment issues at the classroom level. What approaches might teachers bring to bear to enable a strong connection between their instructional emphases for gifted learners and the assessment models that can provide evidence of formative and summative progress among their gifted learners? What tools and strategies might teachers employ to merge their curriculum, instruction, and assessment approaches for advanced learners? And how might students themselves become a part of both determining their assessment process and participating in judging their own learning potential?

The Sulak et al. chapter focused on the use of *curriculum-based assessments* (CBA) to document advanced learning in the classroom. The authors argued that by tying assessment to instruction, educators can better ensure continuous and ongoing instruction that is aligned to each learner's ability. Yet they also cautioned that such assessments have their limitations in respect to technical adequacy. Models of CBA were described in the chapter and examples provided for use in language arts, math, and science. Several of these tools have been used in Javits projects to document pre-post growth in an area of learning such as the Fowler test in science. Others were used to gauge off-level learning such as the MAP test in mathematics. Still others assessed lifelong developmental skills such as writing through assessing the processes that students might have acquired through deliberate instruction. The chapter authors stressed the need for using multiple CBA assessments to evaluate individual student learning through a combination of the measures described or through a portfolio system that integrates the strands of curriculum provided. The chapter concluded with advice for educators on providing professional learning for teachers in using this assessment approach to enhance both teaching and learning.

The Brighton et al. chapter focused on *project-based assessment*, using this alternative assessment approach to capture the comprehensive learnings of gifted students in cognitive and affective areas. The authors outlined the components of project-based tasks and how to implement them in classrooms, providing the context for the use of project-based assessment tools. A comprehensive example was provided of a sample virtual book club project and its accompanying assessment. The authors also provided a process for designing rubrics that is appropriate for use with complex projects, showing how to define the dimensions of the rubric and then how to graduate the expected learning outcomes for a given score from 1 (low) to 5 (high). The authors faithfully acknowledged both the advantages and disadvantages of this approach to assessment, with the positive aspects of connection to the real world, the motivation to learn, and the entry into a profession through mentorships as superior reasons to employ it even as technical adequacy, use of resources, and professional learning needs may be overwhelming. The chapter also addressed implementation efforts for this type of assessment as it is dependent on the merger of curriculum and assessment, which begins with defining expectations and outcomes. This section of the chapter also addressed the resource needs of students to attain success in the model including attention to choice and autonomy, the use of scaffolds, providing feedback, flexible use of hybrid approaches, and connecting across disciplines and to lives.

The Rinn et al. chapter on *self-assessment* examined the types and tools for teaching students how to self-assess important psychosocial skills such as self-esteem, self-regulation, and metacognition as well as core cognitive skills. The authors noted the psychological and educational research supporting the use of the student self-assessment tools of journaling, responding to inquiry about learning experiences, and providing feedback on performance. They reviewed tactics for use in all the core subject areas including a discussion of portfolios, rubrics, learning contracts, and checklists that assist in the process. Self-rating of student products was also discussed in detail, noting the role of rubrics as part of the learning process before, during, and after product development. Sample rubrics were included for educators to consider as they incorporate more student self-assessment into their models of documenting learning. The chapter concluded with a list of recommendations for gifted program coordinators as they implement self-assessment tools in their programs.

Just as assessments need to be tailored in different ways to accommodate gifted learners, so too must they provide accommodations for *special needs learners* within the umbrella of gifted education. The Farah chapter explored the use of alternative assessments for gifted students with disabilities (i.e., twice exceptional) and those who are English language learners. She noted the accommodations and modifications that are needed to optimize these assessment processes with populations of special needs learners. Her chapter provided scenarios, highlighting individual student needs, and described different types of alterations. The chapter concluded with challenges related to accommodations and recommendations.

In examining the scope of learning over time, the next chapter in this section addressed the importance of organizing instruction to promote learning over time, ensuring that cognitive growth and psychosocial growth are both monitored throughout the years of schooling. The Olszewski-Kubilius et al. chapter focused on *learning trajectories* as the "big picture" way to examine gifted student learning. Based on the newest model for examining the talent development process as explicated by these same chapter authors (i.e., Olszewski-Kubilius et al.), the chapter explored the idea of the pathways to competence within domains of study and how they might be enacted through contacts with influential mentors, teachers, and others in a student's inner circle. The chapter was also explicit about the kinds of support that schools might provide to the process, especially in the area of accelerated work. Focused on the importance of ongoing development, the authors stressed the need for critical domain-relevant educational experiences that enable students to transition to the next stage of talent development. They also stressed the need for students to acquire supportive psychosocial skills to enhance the opportunity for advancement to higher stages in a domain.

Issues in the Use of Assessments to Document Gifted Student Learning

Given recent longitudinal studies (see Lubinski & Benbow, 2016) that suggest the way that educational advantage accrues, we can see that learning experiences matter as a continuum of opportunities at different stages of development. While the specifics of the experiences may be important, the fact that gifted students have choices and engage in multiple activities or dosages appears significant to their talent development (Wai & Benbow, 2022). Consequently, the issues associated with assessing gifted students' learning assumes a greater role since students' performance on relevant assessments across pre-K-16 years in school affect the choices they make in their adult lives in respect to advanced education and careers. The more these students have early opportunities, the more they can self-assess their preferences, their aptitudes, and their psychosocial skills to enter certain fields and to live a preferred lifestyle. Perhaps today's students are the first generation to really have these choices as concerns about lack of money, about access to high quality institutions of learning, and about openness to the pursuit of demanding careers has decreased for many of them as scholarships, mentorships, and internships have opened doors. Yet lessons can be learned from these chapters as important issues remain when considering the use of learning assessments for this population.

One lesson to be learned is the need for policymakers to understand the needs and accomplishments of the gifted, often hidden within a precollegiate system that limits their possibilities. Especially for underrepresented groups, latent abilities and skills can adversely affect their futures by not advancing and nurturing their capacities.

A concomitant emergent issue is the lack of data on the actual levels of performance of gifted students in our schools. The Boswell et al. chapter has highlighted this problem. Improving opportunities must be based on evidence of need, not an easy case to be made without the disaggregated state data for this group of learners in our schools.

The topic of many of the chapters in this section of the book explore the question of what types of data are most important to collect that will provide the convincing evidence of growth or the lack of it in gifted learners in advanced programs and outside of them. What evidence is convincing to policymakers at local, state, and even national levels that gifted students are being shortchanged in public schools?

Another issue worthy of discussion is how these assessment data on the gifted might be used and by whom? It is helpful for gifted students and families to have any data that make the case for their level of talent and readiness to learn at the best educational institutions available to them. Yet it is their teachers and mentors who may be the important guides to the next stage of talent development, often not even included in the information loop as these students continue to excel.

Finally, how many assessment approaches should be adopted before assessment drives out instruction in the classrooms of these learners? Teachers only have so much time, so much expertise, and so many resources to devote to any aspect of advanced learning. Is extra time on assessment worth the cost of lost instructional time? This question will continue to be debated as most gifted programs are forced to use the same state assessments in the same way without the option of adding additional assessment data for decision-making on individual students and their programs.

Table II.2 shows the advantages and disadvantages of the tools explored in these chapters to assess gifted student learning through alternative means. Clearly, each approach has its merits, especially in particular contexts and for specific purposes. However, each also has its liabilities in respect to consistency in use and adequacy in technical areas like validity and reliability.

Even as adjunctive tools to existing assessments, alternative assessment models still offer only a limited perspective on the learning progress of these students for next level challenges, given the lack of scope. Portfolio collections over the course of a semester, year, or multiple years have proved successful and may be an option for addressing the issue of scope. Electronic portfolios can be employed to capture the nature and extent of these learnings over time and provide a more accurate picture of advanced learning in schools and classrooms. Other options might be to select two or more such approaches that work well together to provide more data on performance and the potential for future performance. For example, combining product assessments in a domain with the assessment of processes employed to create them would provide a two-dimensional look at two important aspects of creativity. Or, in a second case, combining the reflections of the students themselves on creating a product with the assessment of the product by experts. Or, finally, perhaps employing performance-based assessments in a content domain pre-post alongside student self-assessment as a matched set of evidence bases to support the next level of learning needed to challenge the learner.

TABLE II.2 Advantages and disadvantages of learning assessment tools to judge gifted student advancement

Assessment option	Advantages	Disadvantages	Notes
Traditional state/national achievement tests	– Sound technical adequacy – Well-designed, based on state/national standards	– Designed for the typical learner, not the advanced learner – Ceiling effects – Based on general education curriculum, not advanced	All chapters note these issues.
Program-based assessment (e.g., AP and IB)	– Matched to advanced program outcomes; calibrated to college level classes – Provides off-level complex thinking and problem-solving tasks	– Bound by the program framework – Differs by content area in respect to extent of open-endedness	These assessments are available only at secondary levels where 25 states mandate AP being offered in high schools.
Curriculum-based assessment (performance-based)	– Optimal match to curriculum outcomes and instruction – Provide advanced, open-ended, and higher-level thinking tasks	– May not be used consistently – May not be used pre-post to show growth change – May not have threshold technical adequacy for use beyond one teacher's classroom	Researchers in gifted education for over 30 years have found and/or created performance assessments that do contain necessary technical data. These instruments are underutilized in gifted programs.
Product-based assessment (e.g., projects, problem-based)	– Match to project-based learning – Require assessment of multiple skill sets (process, product, presentation)	– Lack of content validity that relates to content, process, and products – Lack of testing for technical adequacy – Lack of consistency across projects – Lack of collaborative scoring to establish interrater reliability	Evidence of effectiveness of these tools has not been sufficiently studied. Projects tend to originate with individual teachers (or sometimes students), making them idiosyncratic with rubrics that lack technical adequacy.
Self-assessment	– Provide for student voice in the assessment process – Provide important feedback to teachers on instruction – Provide for the development of psychosocial skill sets such as metacognition	– Often lack technical adequacy data – Too limited to be used in isolation – Often not used for specific purposes of improvement of the program/curriculum nor pre-post to assess change	This approach to assessment may be most useful in tandem with performance data on learning to cross-validate findings.

(*continued*)

TABLE II.2 Cont.

Assessment option	Advantages	Disadvantages	Notes
Creativity assessment	– measures growth in domain-specific creativity – requires assessment of person, process, and product – guides future decisions about talent development and career pathways	– lack of experts to assess products and performances – may not have technical adequacy – may be generic and not domain-based – may not use processes authentic to the discipline	Creativity assessments are most useful when used within a domain to determine student's growth. Educators are unaware of technically adequate instruments in this area and use them infrequently.
Altered assessments: Accommodations and modifications	– provides 2E and EL learners with access to assessments – can address variations in performance due to disability or linguistic differences – provides information related to student's true level of performance	– limited guidelines specified in federal or state laws – test alterations are misunderstood by educators and may be misapplied – students may not disclose information about their disabilities	Because accommodations and modifications are highly influenced by each student, research is limited.

Conclusion

The selection and/or development of appropriate and effective learning assessments for use with gifted learners in advanced programs is almost as difficult a task as identification of those learners. As can be seen from the chapters in this section, some of the same issues plague tools to assess learning as impact on selection of students. Instruments may lack ceiling, technical adequacy, and are too narrow in scope to assess broad learning in a domain. Yet there are promising directions in the use of multiple tools that address different aspects of the learning itself. These approaches, taken together, may provide a multidimensional portrait of advanced learning and the learners who benefit from it.

References

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x

Wai, J., & Benbow, C. (2022). Educational interventions on behalf of the gifted: Do they have lasting impacts on development? In J. VanTassel-Baska (Ed.), *Talent development in gifted education: Theory, research, and practice* (pp. 115–131). Routledge.

PART

III

Evaluation of Programs

Introduction to Part III on Evaluation of Programs

Joyce VanTassel-Baska

Introduction to Evaluation

The meaning of evaluation is "to judge or assess the value of an object," in this case the object is a gifted program. The definition of program evaluation is more specific as seen in the *American Psychological Association Handbook of Practical Program Evaluation* (Wholey et al., 2010): "Program evaluation is the application of systematic methods to address questions about program operations and results. It may include ongoing monitoring of a program as well as one-shot studies of program processes or program impact. The approaches used are based on social science research methodologies and professional standards."

Program evaluation is a continuous process that critically examines a program, based on its characteristics, activities, and outcomes. At a formative stage, it involves implementation problems and monitoring of program processes to ensure integrity. At a summative stage, it involves an emphasis on program impact on students, effectiveness of program processes, and cost-benefit analyses. It involves judging the quality of arguments and the credibility of speakers and writers about the merits of the program. A focus is placed on the process being child-centered and therefore focused on learning more than teaching. In many gifted program evaluations, formative data are collected internally and used for short-term improvement of the operation of the program from one year to the next. External evaluations, conducted by experts external to the district, are less frequent, occurring perhaps every 5–7 years.

Models of Evaluation

There are many theories about evaluation and the most important way it should be conducted. Stufflebeam (2001) posited a theory of evaluation that suggested the focus should be goal-based, that one should only judge programs by the standards of what they said they would do. Consequently, the emphasis was on context, input, process, and outcomes of the evaluation. Evaluation questions for assessing gifted programs typically include at least two that fit this model, one related to the processes employed and the other related to student outcomes and other unintended outcomes. Stufflebeam

DOI: 10.4324/9781003285991-22

suggested that evaluators should keep their distance from program implementers, relying on artifacts, observations, and relevant data only to determine findings.

Another evaluation luminary, however, suggested the need to be closer to the work in situ. Stake (1995), therefore, took a more humanistic view of the evaluation process, suggesting that evaluation should be viewed as a client-centered process of interactions with the stakeholders involved with the program to understand better the phenomenon of gifted programs through access to their internal operations and ideas about what was working from the inside out. He devised a case-study approach for evaluation work that cast the evaluator as a part of the evaluation process in a real way, influencing others about how to think about programs and their purposes as well as collecting and analyzing data on multiple facets of program operation.

A third model of evaluation assumed that impact was the critical variable to discern in respect to program implementation (Patton, 2021). Utilization-based evaluation is interested in the outcomes from an evaluation effort, its merits for promoting knowledge, worth, and improvement. In this model, the evaluator is both inside the process to glean an understanding of group and individual perspectives but also outside, collecting large scale data related to outcomes.

Finally, Eisner (1985) pioneered the expertise model where the evaluation process is vested in an expert in the field of study who has the knowledge base to make sense of the data a program produces to support its claims. Judgments would be rendered to determine student and teacher needs, barriers to meeting those needs, opportunities that might be employed to address barriers, and the appropriateness of program goals. This approach is often employed where time is limited, and expertise is available to render a perspective useful to making program decisions.

Often an approach used in gifted program and other educational program evaluations is an accreditation model that assumes that the value of a program may best be seen through its adherence to a field's standards of practice. Thus, some gifted program evaluations have concentrated on this aspect of evaluation, the extent to which the program aligns with best practices in gifted education, especially in the areas of identification, curriculum and instruction, environment, and assessment (Johnsen et al., 2022). Many researchers also use a mixed methods model, combining a view that acknowledges the importance of stakeholder perspectives with the reality of objectively assessing implementation processes.

Purposes of Program Evaluation

So, what is the appropriate way to think about the purposes of an evaluation of programs for the gifted? An evaluation is important to be perceived as:

- A fundamental part of program development, just as critical as identification, or professional learning, or curriculum development;
- A tool for understanding the parameters of program implementation in various contexts, at different levels in different schools and classrooms;
- A process for program improvement in addressing the goals of the gifted program.

These three ways of thinking about evaluation also provide the basis for action planning that may improve the current program but also point to areas that need further development. In that sense the application of evaluation findings, often called "evaluation utilization" demonstrates that evaluation is not just a validation and judgment of current practice but also a guide for future efforts.

The Literature Base on Program Evaluation Results

Very little literature exists on long-term results of gifted program evaluations although several texts provide guidance in how to conduct such evaluations (see Speirs-Neumeister & Burney, 2019; VanTassel-Baska & Feng, 2004). One study (VanTassel-Baska, 2006) examined 20 district evaluations and concluded that gifted programs shared several characteristics: lack of personnel resources to carry out the scope of the program initiatives, lack of focus on program development and expansion, and limited use of differentiation practices. Earlier studies had noted the underutilization of evaluation results (Tomlinson et al., 1994). Recent reports of district emphases in program development report few changes from 20 years ago in the practices being applied, suggesting that newer best practices have not taken hold (Callahan et al., 2017). In a recent national survey that focused on identification and its relationship to services in 293 districts in two states, the National Research Center on the Gifted and Talented (Gubbins et al., 2021) reported that, while schools used domain-specific instruments, they did not provide a separate differentiated curriculum in either math or reading/ELA at the elementary level. While districts in one state were more likely to use cluster grouping, the authors were unsure if any differentiation occurred within the clusters. They also interpreted the survey results to suggest that teachers had control over the curriculum used with gifted learners, noting that critical and creative thinking both were employed.

In recent reviews of curriculum materials, standards' audits, and observed instruction, evaluators have found consistent issues emergent in respect to program implementation in classrooms (VanTassel-Baska & Baska, 2020). Focus groups, interviews, and best practice audits also corroborate these findings (VanTassel-Baska & Brown, 2022; VanTassel-Baska & Hubbard, 2019)). Findings have consistently suggested the need for greater efforts to apply differentiation practices in classrooms, to educate stakeholders on appropriate strategies and materials for use with gifted learners, and to identify more students from underrepresented groups.

A strong focus on program development research in recent years also suggests the need for further growth and development in gifted programs, consonant with research-based best practices (Rogers, 2007). Areas Rogers has highlighted as having the strongest research base include the use of various forms of acceleration, the use of grouping approaches to implement differentiation practices, the use of content-based differentiation strategies, and the use of independent approaches being employed to provide opportunities for gifted learners.

Evaluation Questions

Many evaluations apply the following questions to programs under consideration, questions that address the congruity of goals to implementation, the perceptions of the program by stakeholders, the evidence of best practices in the field of gifted education, and the results for student advanced learning. Table III.1 reflects the questions, the data sources consulted to answer them, and the types of instrumentation used to procure answers.

TABLE III.1 Evaluation questions by data sources and instrumentation

Evaluation question	Data source(s)	Instrumentation
To what extent is the gifted program being implemented according to stated goals?	Document review Classroom observation	Criteria checklist Observation rating scale
To what extent is the program perceived to be effective by relevant stakeholders?	Interviews Focus groups	Interview protocol Focus group protocol
To what extent is the program aligned with best practices in gifted education?	NAGC Programming Standards	Checklist of standards
What data support the effectiveness of the gifted program?	State and national data Student progress data	Quantitative and qualitative assessment tools
What are the strengths and recommendations for improvement?	Analysis of findings from multiple data sources	All of the above

Source: VanTassel-Baska, J., & Baska, A. (2020) *Curriculum planning and instructional design*. Prufrock Press.

Overview of the Evaluation Process

Figure III.1 provides a graphic representation of the total evaluation process. Just as identification and learning progress have their distinctive cycles of implementation, so too does evaluation. The process begins with the questions of interest to a school or district about their gifted program and proceeds to the development of an evaluation design. Next, the evaluator selects or creates instruments needed to answer the questions raised. The design then begins to be activated through the phases of on-site visits to the program by the evaluator as well as off-site work to review program materials, program curriculum, and reports about the program over the past five years. On-site work would include classroom observations, focus groups with stakeholders (i.e., parents, teachers, administrators, and students), and interviews about program issues and concerns. An audit of program operation would follow that examines alignment with national standards in gifted education and implementation fidelity in the classroom. As a result of data collection in these areas, the evaluator provides a report that answers the questions of interest through issuing findings and making recommendations for program improvement. The final aspect of the cycle then is to review the evaluation process itself for areas of improvement, which might include revising surveys that lacked sufficient discrimination or adding an instrument to assess the quality of student products independently of teachers, for example.

The data sources and instrumentation to be used in evaluation are described and critiqued in Chapter 15 that provides readers of this text with important information about carrying out evaluations within districts on an annual basis. Chapter 16 addresses the processes that are being employed in sample school districts to make positive change and improvement in gifted programs, suggesting that it is an ongoing process that requires the use of research-based instruments and processes of data collection and analysis, collaboration across personnel and departments as well as an overriding commitment by the school district leadership to ensure that such improvements continue.

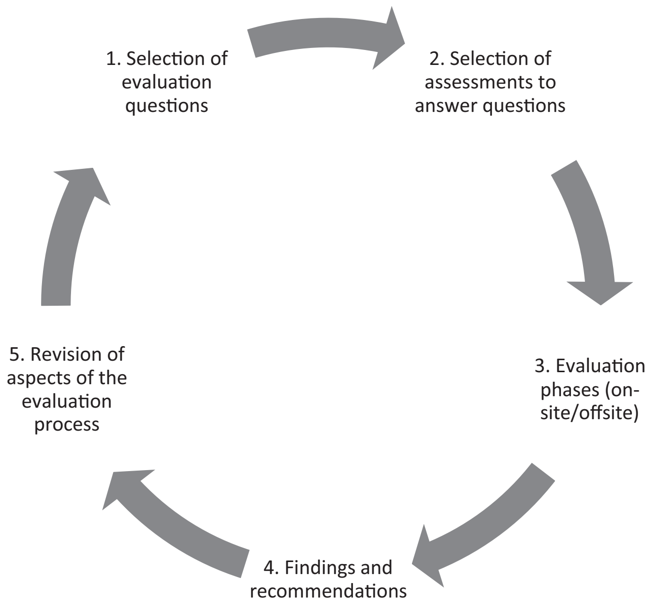

FIGURE III.1 The evaluation cycle in gifted programs and services

Conclusion

The purpose of an evaluation of gifted programs provides an additional service. An evaluation can offer needed credibility for the presence of such programs, help stakeholders understand why they are important, and buttress the argument for gifted education as a legitimate area of education. The field has a substantial research base to support its existence and a trained group of specialists ready to work with students and exemplify best practice. Given that promising mix of variables, it is only fair to suggest that the field should thrive well into the future as evidence of success in gifted programs and services can be documented for all publics who care to examine the data.

References

Callahan, C. M., Moon, T. R., & Oh, S. (2017). Describing the status of programs for the gifted: A call for action. *Journal for the Education of the Gifted*, *40*(1), 20–49. https://doi.org/10.1177/0162353216686215

Eisner, E. (1985). *The art of educational evaluation*. Falmer Press.

Gubbins, E. J., Siegle, D., Ottone-Cross, K., McCoach, D. B., Langley, S. D., Callahan, C. M., Broderson, A.V., & Caughey, M. (2021). Identifying and serving gifted and talented students: Are identification and services connected? *Gifted Child Quarterly*, *65*(2), 115–131. https://doi.org/10.1177/0016986220988308

Johnsen, S. K., Dailey, D., & Cotabish, A. (2022). *NAGC pre-k-grade 12 gifted education programming standards: A guide to planning and implementing quality services for gifted students* (2nd ed.). Routledge.

Patton, M. Q. (2021). *Utilization-focused evaluation*. Sage.

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of research on educational practice. *Gifted Child Quarterly*, *51*(4), 382–396. https://doi.org/10.1177/0016986207306324

Speirs-Neumeister, K., & Burney, V. H. (2019). *Gifted program evaluation: A handbook for administrators and coordinators* (2nd ed.). Prufrock Press.

Stake, R. (1995). *The art of case study research*. Sage.

Stufflebeam, D. (2001). Evaluation models. In G. T. Gary & J. C. Greene (Eds.), *New directions for evaluation* (pp. 7–99). Jossey-Bass.

Tomlinson, C., Bland, L., Moon, T., & Callahan, C. (1994). Case studies of evaluation utilization in gifted education. *American Journal of Evaluation*, *15*(2), 153–168. https://doi.org/10.1177/109821409401500205

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 gifted programs: A clarion call for enhanced gifted program development. *Gifted Child Quarterly*, *50*(3), 199–215. https://doi.org/10.1177/001698620605000302

VanTassel-Baska, J., & Baska, A. (2020) *Curriculum planning and instructional design*. Prufrock Press.

VanTassel-Baska, J., & Brown, E. (2022). An analysis of stakeholder perceptions of gifted programs: A report card on gifted program performance**.** *Gifted Child Today*, *45*(3).

VanTassel-Baska, J., & Feng, A. (2004) *Evaluation of gifted programs*. Prufrock Press.

VanTassel-Baska, J., & Hubbard, G. (2019). A review of the national gifted standards implementation in eight districts: An uneven picture of practice. *Gifted Child Today*, *42*(4), 215–228. https://doi.org/10.1177/1076217519862336

Wholey J., Hatry, H., & Newcomer, K. (2010). *APA Handbook of practical program evaluation*. Jossey-Bass.

CHAPTER

15

Evaluation Tools to Assess Gifted Programs

Selection and Use

Cheryll M. Adams and Melanie Caughey

This chapter builds on the foundational information on program evaluation in the introduction to this section to answer the question "How do I choose appropriate tools (any protocol, instrument, scale, assessment, and such that might be used) to conduct my program evaluation?" We will begin our response with background on program evaluation and outline criteria for selecting effective evaluation tools and use of the National Association for Gifted Children (NAGC, 2019) *Pre-K to Grade 12 Gifted Programming Standards* to provide direction for the selection and use of such evaluation tools. Then we describe the different evaluation tools that practitioners can use, including how each tool meets the NAGC criteria and how they might be interpreted and used effectively to improve programs and services, particularly for diverse learners.

Background

Program evaluation is a formal process during which practitioners collect data to determine two equally important outcomes: (a) if a program is or is not meeting its goals, and (b) why it may or may not be meeting its goals (Callahan et al., 2017). Evaluations may be conducted internally by qualified school or district personnel or externally by experts in the field of gifted education. The scope of the evaluation may focus on one area or component of the program at a time to make the process more manageable for the practitioner. External evaluations conducted by experts in evaluating gifted programs generally examine multiple components of the program and may include more than one district. It is recommended that internal evaluations be conducted annually and external ones every five to seven years. Program evaluations must be frequent and ongoing for stakeholders to get meaningful results.

To evaluate a program, the components of the program should be clearly documented with a plan for the program evaluation in place. According to a survey conducted by Callahan et al. (2017), less than 50% of 1566 school district personnel responsible for gifted education indicated they had a program evaluation component in their gifted

DOI: 10.4324/9781003285991-23

plans to monitor programs and services for continuous improvement. Most districts reported limited internal evaluations with the evaluators being educators from within the gifted program. In the most recent *2018–2019 State of the States in Gifted Education* report (Rinn et al., 2020), the data indicated that less than half of states require local education agencies' (LEA) gifted plans to be submitted and/or approved. While 20 states included program evaluation as a required component in the plan, only three states had evidence on their Department of Education websites that either the state education agencies (SEAs) or LEAs had completed any program evaluations.

Furthermore, there was not clear information on the tools that practitioners used to conduct these evaluations. Fifteen state plans had evidence that LEAs were required to complete a self-assessment checklist or rating scale, generally for eligibility to receive state funds. Thirteen of those states provided actual evaluation tools for LEAs to use to conduct their own evaluations. These tools included forms to document the number of gifted students in the program and their demographics; student, parent, and teacher surveys; and self-assessment checklists and rating scales. On a cautionary note, however, no reliability or validity information was found on any of the program instruments developed by SEAs or LEAs and posted on their websites. This finding aligned with earlier study results from Callahan et al. (1995). Clearly, there continues to be a lack of reliable instruments with validity evidence being used to evaluate gifted programs by SEAs and LEAs.

Criteria for Effective Evaluation Tools

Practitioners must have appropriate tools to gather information that will provide data to indicate areas of the program or services that are effective at improving student performance and those that are not. Although there are many tools used for this purpose, not all tools are effective due to lack of evidence for validity and reliability and other psychometric properties and a lack of use in prior studies.

Evidence-based practices in Standard 5 of the NAGC Programming Standards provide direction for the use of assessment instruments and tools in program evaluation (NAGC, 2019). They highlight the importance of (a) using reliable and valid instruments, (b) using assessments that allow gifted learners to show adequate yearly progress, and (c) ensuring fidelity of implementation in all components of gifted education programming that influences student outcome. Fidelity of implementation describes how well the delivery of an intervention follows a protocol or program model as it was intended (e.g., O'Donnell, 2008; Lakin & Rambo-Hernandez, 2019). For example, Standard 5.8.2 states practitioners should be aware of how fidelity of implementation influences student outcomes; hence, using a high-quality classroom observation scale can assist in determining whether differentiation of instruction is occurring in the classroom at an appropriate level to ensure that gifted learners are challenged (VanTassel-Baska et al., 2004). Similarly, using a rubric for rating exemplary curriculum for gifted learners can pinpoint areas of the curriculum that might need additional content or revision of existing content (VanTassel-Baska & Baska, 2019).

To ensure practitioners have a means of judging the quality of the evaluation tools they might use, we reviewed literature on program evaluation (e.g., Callahan, 2009, 2017, 2020; Speirs Neumeister & Burney, 2019; VanTassel-Baska, 2006), the NAGC standards (e.g., Cotabish et al., 2017, 2020; NAGC, 2010, 2019) and evaluation tools (e.g., Callahan et al., 1995; Farah & Chandler, 2018; National Research Center on the Gifted and Talented, n.d.-a, n.d.-b) and considered the purpose of the tools, the use of the tools with gifted students, the interpretation of the data, the potential for bias, the research-based evidence, and the psychometric properties of the tools. Additionally, we provided guidance for choosing reliable and valid data sources. Examples are provided in the Resources section and Appendix D.

Selecting Appropriate Evaluation Tools

Searching for Tools

We searched the gifted education literature to find references to tools that would be appropriate to use for gathering data when conducting program evaluations. The types of evaluation tools we selected for this chapter include surveys, focus groups, interviews, rating scales, observation protocols, document reviews, and templates. We provided an overview of these tools and created both a list (see Resources) and a chart with specific examples of each tool (see Appendix D) that researchers have used to evaluate programs. We hope that practitioners will find this information useful when they need to choose appropriate tools to conduct their own program evaluations.

While most of these tools were not intentionally aligned with either version of the *PreK-12 Gifted Programming Standards* (NAGC, 2010, 2019), aligning a tool to the NAGC standards would not be a difficult task for a practitioner with a working knowledge of these standards. It is recommended that practitioners do that as part of their program development work prior to evaluation.

Data Sources

To obtain reliable data that is valid for its purpose, the sources that are used must be credible and useful. Two necessary sources of information are stakeholders and tools (instruments).

Stakeholders

Common stakeholders in schools include administrators, gifted coordinators, teachers who work within the gifted program, teachers in general education classroom, school counselors, parents, and students. In some districts, there may be a gifted program advisory committee with additional members drawn from the community, nearby institutions of higher education, the school board, and/or central office personnel. It is vital that stakeholders represent the diversity of the school or district regarding gender, culture, ethnicity, language, and socioeconomic status (Mun et al., 2020; Robbins, 2019).

Stakeholders' knowledge and perceptions of gifted students and the gifted program may vary widely from those who have advanced degrees in gifted education to those who have little background in the field. Some stakeholders will have negative perceptions of gifted learners and their needs while others will perceive them more positively. To obtain credible and useful data, whoever is conducting the evaluation must determine which stakeholders should gather the data needed for the selected tool. Clearly, parents should not use an observation scale to determine if the teacher is implementing a particular curriculum model with fidelity. Similarly, students should not be asked to conduct a document review of program components. Data from inappropriate stakeholders will not help practitioners make informed decisions about the program. See Robbins

(2019) for a comprehensive discussion of gifted education stakeholders' strengths and potential barriers.

Tools

Practitioners should collect both quantitative and qualitative data using tools that align with the NAGC Programming Standards and the goals of the program (Callahan, 2009, 2017; Robbins, 2019; Speirs Neumeister & Burney, 2019). This allows the responses gathered to illuminate specific components of the program so that any recommendations for change are based on best practices (Callahan 2009, 2020).

Quantitative data should be obtained by using tools that have psychometric properties, such as reliability, validity, and norming data. Standardized measures of achievement and ability and some of the classroom observation protocols, rating scales, and surveys we found in our search have psychometric properties. Psychometrics are discussed further in the Interpretation section.

Tools like interviews, focus groups, and some surveys provide qualitative data (e.g., descriptions, opinions, and perceptions). For these tools, the designer must be cognizant of the information that needs to be collected to be able to construct good open-ended questions that produce results valid for the evaluation's purpose. These narrative data are then analyzed through content analysis (Patton, 2015). Tools like observation protocols, document reviews, and rating scales should require users to have specialized training, knowledge, and/or skills to ensure the results are reliable.

Based on the results of our search for tools and the examination of the literature on program evaluation, we analyzed seven types of tools commonly used in evaluating gifted programs by considering their characteristics, purpose, psychometric properties, use with gifted students, use with diverse students, ease of interpretation, and bias.

Surveys. One of the most common types of tools used in program evaluations is surveys, which are designed to gather opinions, perceptions, and other information from stakeholder groups. The practitioner must first decide which stakeholders will be the best sources to provide the data needed. Then, they will design questions so that the respondents' answers will reflect the perceptions of different stakeholder groups. These questions may require forced-choice responses (e.g., asking respondents to select an answer from the choices presented) or open-ended responses (e.g., asking respondents for explanations). Hence, surveys can gather both quantitative and qualitative data. Asking the same questions to different stakeholder groups (e.g., parents, teachers, and students) and disaggregating the data provides different perceptions of the same program areas; thus, it is important to gather demographic data.

When constructing a survey, respondents seem to prefer shorter surveys or those that do not involve lengthy responses to many open-ended questions (Feng, 2004; Speirs Neumeister & Burney, 2019). Understanding respondents' perceived preferences for brevity and providing clear, simple directions help to increase the response rate. Be sure to avoid jargon and to provide space in case someone would like to elaborate on an answer.

Focus Groups. A means of gathering qualitative data, focus groups are small-group interviews of specific stakeholders conducted by a facilitator using a protocol of structured questions. A knowledgeable facilitator with good interview skills is needed to ask the questions, as they need to know how to avoid cueing participants for particular responses, to follow up on a line of questioning when necessary, and to watch for group think and the participants' body language. Facilitators should be clear on the outcome (e.g., what they want to learn from a group) beforehand to ensure all the questions are going to answer the questions appropriately. Questions should be open-ended, with follow-up questions included in the protocols in case the facilitator needs to gather additional information.

The group dynamics of the participants provide context for interpreting the data gathered (Feng & Brown, 2004; Krueger & Casey, 2001; Speirs Neumeister & Burney, 2019). Generally, there are six to ten participants (Speirs Neumeister & Burney, 2019) to allow for diverse opinions and opportunities for everyone to share while seated in a circle. The participants should be chosen because they have experience with or knowledge of the information that needs to be gathered, reflect the diversity of the stakeholder group, and are at similar administrative levels (e.g., do not include department chairs with teachers they supervise; Feng & Brown, 2004; Krueger & Casey, 2001). Be careful to avoid participants who have an agenda and only speaking with those who are interested, as these are factors that will impact the usefulness of the data. However, it is important that the participants are comfortable speaking with each other so that everyone can speak freely.

Interviews. Like focus groups, individual interviews involve meeting with stakeholders face-to-face. Other than demographics, the data gathered are typically qualitative in nature. As with the facilitator of focus groups, the interviewer must be knowledgeable about the topic and know when to ask for further detail, be clear on the outcome, be careful not to cue responses, and ask open-ended questions.

The individuals chosen to participate in interviews must have enough knowledge to respond completely to the questions asked. Because only the interviewer and interviewee are in the room, people may be more willing to provide a response that reflects their true beliefs without the fear of possible consequences later.

Interview questions are often used to gather data about curriculum, instruction, and program perceptions (Callahan, 2009). Speirs Neumeister and Burney (2019) suggest including interviews at the beginning of a program evaluation after the evaluation team is familiar with any available published documents and desires clarification. Subsequent follow-up interviews may occur later if there is a need for additional information. Questions should only contain one idea and be free of jargon (Callahan, 2009; Fink, 2003).

Rating Scales. Rating scales are evaluation instruments that ask raters or respondents to make judgments about the ideas or descriptions presented in the scale; they are not checklists (Acar et al., 2016; Spector, 1992). We will limit our discussion to those used in evaluating gifted program components.

The person completing the rating scale is the data source. Respondents are asked to indicate whether they agree or disagree and to what degree. There are commonly three to seven of these frequency statements, and the language used reflects what is being evaluated. For example, the My Class Activities scale (Gentry & Gable, 2001; Pereira et al., 2010) uses a five-point frequency scale (never, seldom, sometimes, often, and always). A drawback of rating scales is that respondents may choose the rating they perceive the practitioner wants or does not want (i.e., response bias) rather than the rating reflecting their true belief (Spector, 1992).

The technical qualities of the rating scale should be assessed before using it. Additionally, Pereira et al. (2010) caution that an instrument used with one population (e.g., general education students) needs to be re-normed when using it with another population (e.g., gifted students in an enrichment program). When the rating scale is used to assess a program (Cotabish, et al., 2020) or a product such as a curriculum unit (Beasley et al., 2017; Purcell et al., 2002), the rater must have the expert knowledge necessary to accomplish the task. Before selecting a scale for use, practitioners should make sure there are clear criteria to delineate between the different scores, raters/evaluators are familiar with the purpose of the instrument and the language used within it, and that they can provide any necessary training to ensure appropriate use of the instrument.

Observation Protocols. To assess the classroom interactions between gifted students and their teachers, it is recommended that practitioners use an observation protocol to provide a structured way to gather the information. There are multiple observation protocols that have been used in research studies with gifted students (e.g., Farah & Chandler, 2018) for a practitioner to either choose or to use as a model for creating a new tool. Regardless of the tool used, it is crucial that the practitioner ensures that it is aligned with the professional standards (Farah & Chandler, 2018; Peters & Gates, 2010). While many of the protocols used in gifted education research were created before the NAGC and Council for Exceptional Children, The Association for the Gifted (CEC-TAG) issued the teacher preparation standards (NAGC & CEC-TAG, 2013) and prior to the revised NAGC Programming Standards (2019), some experts are updating and/or re-validating the tools for current use (e.g., Johnsen et al., 2020; Pereira et al., 2019, 2021; VanTassel-Baska et al., 2020). Farah and Chandler (2018) also recommended selecting a protocol that meets the following criteria established by the American Psychological Association (2014): (a) standardized administration procedures, such as training, procedures, and scoring; (b) consistency, meaning if the tool is reliable within different settings, raters, observation periods, etc.; and (c) association with student outcomes, chiefly validity evidence linking the tool to achievement scores.

Practitioners must also make sure they select and/or create a tool that best fits their needs. First, they should ensure that the tool aligns with the services being provided. For example, there are multiple instruments that are intended to measure how teachers differentiate instruction (e.g., Avery et al., 1997; Cassady et al., 2004; Johnsen et al., 2002, 2013, 2020; VanTassel-Baska et al., 2006, 2020; Westberg et al., 1993) while a protocol that focuses on the teacher may work better with different service delivery models (e.g., Hansen & Feldhusen, 1994; Peters & Gates, 2010). Second, practitioners should consider how much time and availability they have to collect the necessary information. Protocol data can be collected in a variety of ways, including teacher self-report (Pereira et al., 2019, 2021), single-class observations (e.g., Avery et.al., 1997; Peters & Gates, 2010; VanTassel-Baska et al., 2006, 2020) and observations of targeted students (e.g., Cassady et al., 2004; Tomlinson et al., 1995; Westberg et al., 1993). VanTassel-Baska (2004) suggests all program evaluations should include observation tools to substantiate "program document claims and stakeholder perceptions" (p. 105).

Document Reviews. Practitioners should also collect documents to support and supplement the data they gather from stakeholders. Types of documents include identification procedures, curricula and lesson plans, staff development and training materials (Callahan, 2009), other relevant school policies, agendas and minutes from meetings, school and student data, and emails. In other words, any document that contains information on a school's gifted program can provide useful evidence to help answer any outstanding questions and to better understand the context of the school and program (Newcomer et al., 2015). Some of the documents may help evaluators complete rating scales or templates (Callahan, 2009). For example, Cotabish et al. (2017) explain how an example school used student demographic information, standardized test scores, and parent-teacher conference sign-in sheets as evidence that they were meeting the NAGC Programming Standards.

When analyzing documents, it is important to consider both the target audience and the purpose of the information. In addition, practitioners should consider what the documents do not include, as missing information can provide perspective on the state of the program (Bowen, 2009). Checklists can help practitioners determine what information is present and what information is missing for documents collected like curriculum (e.g., VanTassel-Baska & Baska, 2019) or even the overall program (Maryland Department of Education, 2021; Virginia Department of Education, 2013).

Templates. Practitioners should also consider using templates to help them gather information to help them review their programming and services. These are often created based on some type of performance indicators (Newcomer et al., 2015). When evaluating a gifted program, it makes sense to create or use a template aligned with the NAGC Programming Standards like the Gap Analysis Chart in Cotabish et al. (2017). Gap analysis charts have also been created to help schools evaluate their programs based on their state standards (see Resources for an example). Unlike rating scales, surveys, and observation protocols, a template does not have any psychometric properties, though it should include space for concrete evidence to show how the criteria are met and allow for multiple stakeholders to contribute to the document to ensure some measure of validity and reliability.

Practitioners need to consider the kind of data they need to organize and the most sensible way to organize that data so others can understand what they are examining before creating and/or selecting a template for use.

Interpretation

We discuss the interpretation of data gathered by the tools here to avoid the redundancy of providing the same information under each tool.

Quantitative Data

Numerical or scaled (quantitative) data (e.g., demographics) can be analyzed using descriptive statistics (e.g., mean, median, mode). Practitioners should also determine whether a tool is reliable and valid for its purpose, providing guidance on how to interpret the data.

Rating scales, surveys, and observation protocols should meet criteria for reliability and validity (Callahan 2009; Pereira et al., 2010, 2021; VanTassel-Baska, 2006; VanTassel-Baska & Baska, 2019). When using a rating scale, check to see if norms have been provided for determining how the ratings compare to the ratings of the national population. Norms are used to compare the score of an individual to the distribution of scores in a population (Spector, 1992). When possible, the norming population should reflect the population of the school. If not, then local norms may need to be created. (See Peters and McBee [2019] who have created a step-by-step guide for creating local norms.) If a practitioner is using a scale to rate an aspect of a program or a product such as a curriculum unit, measures of content validity, construct validity, criterion-related validity, internal consistency, and inter-rater reliability are important (Beasley et al., 2017; Purcell et al., 2002). Content validity ensures that the items are measuring the content as intended; construct validity ensures the items are measuring the hypothetical constructs that are being studied; and criterion-related ensures that the rating scale predicts the effectiveness of the unit (Creswell & Creswell, 2018). Inter-rater reliability will help to show that different raters can use the scale/protocol and get similar results. Researchers may also report the internal consistency of their tool, which indicates if the items are assessing the same construct. An optimal Cronbach's alpha should be between 0.70 and 0.99 (Creswell & Creswell, 2018). Lastly, some tools may have user manuals that provide guidelines for analyzing the data to consult before use (see VanTassel-Baska et al., 2005).

It is also important to consider item bias, which Ryser (2018a) explains is social (strengthens stereotypes) and/or statistical. Social item bias occurs when groups will score differently on an instrument based on their background and not what is purported to be measured by the instrument. Statistical item bias occurs when different groups have different probabilities of receiving a score based on differential item functioning (DIF). Both types of bias can be resolved by having experts review the instrument for any stereotypes and assumptions that may be offensive toward a particular group (Ryser, 2018a).

Qualitative Data

Written (qualitative) responses should be categorized/coded. Content and pattern analysis are used to analyze qualitative data, and the resulting themes help to triangulate data gathered using other tools (Callahan, 2009; Feng & Brown, 2004; Krueger & Casey, 2001; Speirs Neumeister & Burney, 2019). Because of the personal nature and the opportunity to probe answers in real time, qualitative data can be quite powerful.

Reliability and validity are also important when working with qualitative data but do not involve statistics. Reliability is about having a consistent approach despite the presence of different researchers or among different projects, while validity is about checking the accuracy of the findings (Gibbs, 2007, as cited in Creswell & Creswell, 2018). For reliability, having multiple individuals working on the analysis to establish common patterns and themes is important. For validity, multiple strategies like triangulation of sources, member checking, discrepant information, and external auditors can help provide more accuracy (Creswell & Creswell, 2018). Ryser (2018b) explains qualitative reliability as (a) consistency (e.g., are all samples collected and/or observations made in the same way?), (b) stability/reliability (e.g., are there differences in the samples collected related to the time period?), and (c) scorer (e.g., are there differences in agreement between different observers?). She also connected quantitative concepts of validity to qualitative research: (a) content-oriented (e.g., are the samples collected and/or observations made "an adequate representation of what is being measured?"), (b) criterion-related (e.g., are the samples collected and/or observations made related to what is being measured?), and (c) construct (e.g., are the tools used based on theories and models to ensure they are measuring what you are looking for?) (Ryser, 2018b, p. 53).

Fowler (2014) underscores the need to design questions that respondents will answer consistently and reflect what the practitioner asks as a means of providing qualitative reliability and validity. Writing questions that are free of cultural and linguistic bias, concise, free of educational jargon, contain no emotional trigger words, and have only one aspect to consider can assist with this and simplify data interpretation (Callahan, 2009; Feng, 2004; Mun et al., 2020; Speirs Neumeister & Burney, 2019). Potential sources of bias in group settings include non-neutral questions, participants with a particular agenda, facilitator cues, body language of participants, and group think (Feng & Brown, 2004; Krueger & Casey, 2001; Speirs Neumeister & Burney, 2019).

Use with Diverse Populations

The issues of identifying racially, culturally, ethnically, and linguistically different (RCELD) learners for gifted programming, as well as those who are twice-exceptional (2e), are discussed in depth in previous chapters in this book. We are focusing on problems inherent in the instruments themselves. Within discussions of excellence, equity and diversity are "race, culture, class, ethnicity, income, gender, sexual orientation, linguistic differences, and learning differences" (Council for Exceptional Children, The Association for the Gifted, 2020, p. 4). Baldwin (2002) stressed that "[t]he evaluation of the success of the program designed for gifted students from diverse cultural backgrounds depends on what types of evaluative tools are selected" (p. 146). Therefore, all program evaluations should be designed using a lens situated in diversity and equity so that the participants and tools used in the evaluation reflect the diversity of the school or district.

Participants

For responses to parent surveys and rating scales to accurately reflect the diversity of the school, the tools used to gather the data may need to be translated into other languages, include postage paid return envelopes, and be edited so they minimize cultural bias. Student versions of these instruments must be tailored to the grade level of the children involved, making sure all students can participate (e.g., large print for seeing-impaired learners). Language should be easily understandable regardless of culture, language, race, or ethnicity. Participants in interviews and focus groups, whether adults or students, should be selected to reflect the diversity of the school, and interpreters need to be provided for those who may be hearing impaired or speakers of other languages.

Implicit Bias

Practitioners must ensure that all who are using observation protocols understand how diverse gifted students may respond differently to instruction and activities in the classroom. Similarly, they must be aware of implicit bias towards gifted teachers of color, no matter how unintended (Grissom & Redding, 2017). According to Davis (2019), there is a "cultural mismatch between our predominantly White, female, middle-class teaching force and our increasingly culturally diverse population of students" (p. 52). Hence, reviews of curriculum should make certain that gifted RCELD and 2e learners have opportunities to see themselves reflected in the curriculum, instruction, and literature of all subjects (Ford et al., 2018). For example, Ford (2009) created a Multicultural Curriculum Checklist, which includes questions such as (a) "Were students given opportunities to examine their own biases (stereotypes and assumptions)?" (b) "Were materials and resources multicultural?" and (c) "Were a variety of assessment or evaluation strategies adopted/used?"

All gifted program documents should be examined using criteria like these to determine areas that do not address the needs of RCLED and 2e learners.

Putting It Together

The following section is an example of an internal evaluation where practitioners are determining the extent to which teachers are differentiating instruction for gifted learners. See Johnsen et al. (2020) for an in-depth example of a district-wide program evaluation and VanTassel-Baska (2006, 2019) for an example of program evaluation across multiple districts.

A Hypothetical Example of Internal Evaluation at the District Level

Oakwood Public Schools (OPS) is an above average, public school district located in a medium-sized city, situated in a county that includes urban, suburban, and rural schools. The student population of OPS is 42% White, 34% Black, 12% Hispanic/Latinx, 5% Asian, and 7% Multiracial. Sixty-seven percent of the students are considered economically disadvantaged. Gifted students represent 8% of the district's school population and closely reflect the diversity of the district.

Dr. Azari, the superintendent, ensures that the district completes an annual internal evaluation of its gifted services, examining one or more components of the program. Every five years, she hires a nationally recognized expert in gifted education to conduct an external evaluation on a broader scale. Thus, every year an evaluation report is presented to the district's broad-based planning committee and each building level advisory board. Following that, the report is presented at a school board meeting and posted on each school's website prior to sending it to the state director of gifted programs for inclusion on the state website.

The district's Gifted Education Coordinator, Dr. Palmer, regularly provides professional learning opportunities for all teachers who have gifted students in their classrooms. The gifted program is a cluster-group model with five to seven gifted learners in the general education classroom. Three years ago, her focus was on differentiating instruction in the general education classroom. Experts in this area presented a series of workshops targeting specific strategies to assist teachers with differentiation, specifically for gifted learners.

It is time for Dr. Palmer to conduct the yearly internal evaluation of the district's gifted program, and she has chosen this year to focus on differentiation in the cluster classrooms in the 15 elementary schools. Specifically, she wants to determine the extent to which teachers are adapting instruction to learner differences in these classrooms and if parents, students, and teachers feel students are being challenged. She has had experience using the research-based Classroom Instructional Practices Scale (CIPS) which consists of four areas: content (i.e., what students need to know, understand, and be able to do), rate (i.e., pace of instruction), preference (i.e., most effective ways to learn the content), and environment (i.e., room arrangement to foster cooperation and collaboration). The learning progressions within each area move from the least adaptative to the most adaptive practices for meeting learners' varying needs. Trained observers rate a teacher's classroom practices using data that have been gathered from classroom observation of a learning task, interviews, questioning, and room arrangement (Johnsen, 2016; Johnsen et al., 2020).

She selects three research-based Program Satisfaction Surveys (see Teacher, Parent, and Student in the Survey Resources; Delcourt & Evans, 1994b, 1994d, 1994e) that will assist with triangulating data. All of the instruments that will be used in the internal evaluation are reliable and valid for the purposes of this internal evaluation. Because nearly 60% of the student population is non-White and nearly 70% are considered economically disadvantaged, Dr. Palmer chose instruments that have been used with RCLED populations (see Resources for more information on these instruments).

To begin, Dr. Palmer met with the Gifted Resource Teachers (GRT) from each of the 15 schools in the district for a week-long training session in the use of the CIPS, including the classroom observation instrument (i.e., observation of a learning task, teacher and student questions, and room arrangement) and interview protocol. The GRTs then scheduled their classroom observations and interviews at times convenient for the teachers' schedules. Surveys were distributed to teachers and students prior to the observation and collected by the GRT. Each GRT sent a survey and self-addressed stamped envelope to the parents/caregivers of all the gifted students who were placed in the cluster classrooms.

Data collected from all sources were analyzed and triangulated to provide information about the extent to which teachers were adapting instruction to learner differences in these classrooms. For example, all three surveys contain a question that about the work being appropriately challenging for the students. The observation instrument lists features that suggest choice and challenge (e.g., methodology of the discipline, compacting, work beyond grade level), and several of the rankings on the CIPS describe features of more challenging work (e.g., enrichment, acceleration, creative and critical thinking, higher level questioning).

Analysis of the CIPS data demonstrated that 60% of the cluster teachers were using content based on broad-based themes and authentic methods, post-testing at set times and allowing acceleration and enrichment, allowing variation in tasks, and arranging the room for student interaction with interest centers present, indicating teachers had a good understanding of differentiation, although there was still room for growth based on the dimensions of the scale. Ten percent of the teachers demonstrated exemplary differentiation in content, process, product, and learning environment with rankings at the "most adaptative practice" level on all four areas of the CIPS. Thirty percent of teachers were still providing the same work to all students, with differentiation only by reading level, in traditional classroom arrangements. Triangulation of the CIPS data with the surveys provided support for these findings. On closer inspection, Dr. Palmer discovered the teachers who were providing little differentiation were not in the district when the differentiation workshops were held.

As a result of the internal program evaluation, Dr. Palmer developed a multi-faceted plan for improving teachers' ability to differentiate instruction for gifted students. She developed workshops targeting the knowledge, skills, and understandings necessary to help the teachers who were differentiating instruction quite well to become better. GRTs scheduled opportunities for these teachers to observe those who were excellent at differentiation. For those who had little background in differentiation, she provided professional learning opportunities including workshops, book studies, and mentoring. She encouraged those teachers whose classrooms were models of exemplary differentiation to submit proposals to local, state, and national conferences to share their ideas and learn from colleagues outside the district.

Although the program evaluation results drove these initial positive changes in teacher practices and expectations, the district saw the need for continuing the momentum by developing a timeline and a list of priorities to ensure that changes in practice resulted in changes in learner outcomes and teacher beliefs and attitudes (Guskey, 2002). Having supportive administrators at the district and building levels helped circumvent some problems that are often roadblocks to implementing change (e.g., resources, time).

After analyzing the evaluation report, these administrators were convinced of the importance of differentiation and realized that the increases in the expertise of the cluster teachers meant all students in the class would benefit. District funds garnered through the state's ESSA plan were then earmarked to provide district-wide continued professional learning in the area of differentiation to effect increased learner achievement outcomes in an effort to decrease the achievement gap, particularly for gifted learners. Administrators examined current teaching schedules in the elementary schools to determine what changes could be made to future schedules to allow for more opportunities for peer coaching, collaborative teaching, and common planning times which would support the differentiation initiative. Monies to provide necessary resources for cluster teachers to use to differentiate became available by prioritizing the initiative. The GRTs monitored the changes and provided support as needed. The administrators anticipated that after five years the changes in teachers' ability to differentiate would result in positive learner outcomes, particularly for gifted learners, and the beliefs and attitudes of those who were initially skeptical of the practice would be changed.

Conclusion

Program evaluation is a vital component of gifted programs and is necessary in providing frequent and ongoing evidence that gifted students are benefitting from the program. The NAGC Programming Standards (2019) underscore the need for using reliable and valid assessment instruments and tools, providing time and resources to conduct program evaluations, creating and implementing purposeful plans, and disseminating results. Despite these standards, we found little indication that SEAs or LEAs have conducted program evaluations or disseminated the results. Furthermore, evidence in state and local documents point to a lack of psychometrically sound tools being suggested for use in program evaluations. Thus, practitioners need access to and guidance in selecting readily available tools that are both valid and reliable for evaluating their programs.

Moreover, program evaluations must be designed and conducted using instruments that minimize bias and use a lens of diversity if the results of the evaluation are going to be usable and reflect the experiences of RCELD and 2e learners. Ignoring issues of implicit and explicit bias in stakeholders, evaluators and instruments invalidates the data gathered. Thus, specific examples of psychometrically sound tools that have been used with diverse populations have been identified in Appendix D.

Our initial question at the beginning of the chapter was "How do I choose appropriate tools to conduct my program evaluation?" By using the descriptions of various types of common tools used in program evaluations, analyses of specific tools, guidance for ensuring reliable and valid results, and the foundational lens of diversity, practitioners should now be able to answer this question as they plan their own program evaluations.

Recommendations

1. Program evaluation should be a required component of all gifted program plans at the state and local levels and comprise more than a checklist of completed tasks and processes.

2. All SEAs/LEAs should ensure that the time and money necessary to conduct annual program evaluations internally and every five years externally are available.
3. Program evaluations should be aligned with the revised NAGC Programming Standards (2019).
4. Both qualitative (interview and focus groups) and quantitative (student data, scales, surveys, and other protocols) data should be collected.
5. *Reliable* tools that are *valid* for the evaluation's purpose should be selected, based on the specific parameters of each program evaluation and the diversity of the population.
6. Individuals conducting program evaluations must have the necessary qualifications and training to do so (i.e., at least 12 hours of university hours of instruction in gifted education); consider including the school psychologist and/or another individual with a background in assessment data analysis to be a part of the local team for annual evaluations.
7. All pertinent stakeholders should be included in any program evaluation.
8. Results of program evaluations should be disseminated at the local and state levels and be easily accessible on school and state websites.

Resources

Evaluation Guides

Callahan, C. M. (2009). Evaluation for decision-making: The practitioner's guide to program evaluation. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted* (pp. 119–142). Prufrock Press.

Callahan, C. M. (2017). Developing a plan for evaluating services provided to gifted students. In R. D. Eckert & J. H. Robins (Eds.), *Designing services and programs for high-ability learners: A guidebook for gifted education* (2nd ed., pp. 225–238). Corwin.

Callahan, C. M. (2020). Gifted programming standards: Understanding their role in program design and evaluation. In J. H. Robins, J. L. Jolly, F. A. Karnes, & S. M. Bean (Eds.), *Methods & materials for teaching the gifted* (5th ed., pp. 127–144). Prufrock Press.

Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instrument and evaluation designs used in gifted programs* (Research Monograph 95132). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95132.pdf

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation* (4th ed.). Wiley.

Speirs Neumeister, K., & Burney, V. H. (2019). *Gifted program evaluation: A handbook for administrators and coordinators* (2nd ed.). Prufrock.

VanTassel-Baska, J., & Feng, A. X. (Eds.). (2004). *Designing and utilizing evaluation for gifted program improvement*. Prufrock Press.

Evaluation Tools

National Research Center on the Gifted and Talented. (n.d.-a). *NRC/GT Instrument Bank Instruments*. University of Connecticut. https://nrcgt.uconn.edu/research-based_resources/ibcd05/

National Research Center on the Gifted and Talented. (n.d.-b). *NRC/GT Instrument Bank Instruments II*. University of Connecticut. https://nrcgt.uconn.edu/research-based_resources/ibcd09/

Surveys

Delcourt, M. A. B., & Evans, K. (1994a). *Administrator satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Administrator-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & Evans, K. (1994b). *Parent satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Parent-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & Evans, K. (1994c). *Qualitative extension of the learning outcomes study* (Research Monograph 94110). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm94110.pdf

Delcourt, M. A. B., & Evans, K. (1994d). *Student satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Student-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & Evans, K. (1994e). *Teacher satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Teacher-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & McIntire, J. A. (1993). An investigation of student learning outcomes: Results of a program satisfaction survey. *Newsletter of The National Research Center on the Gifted and Talented, Winter*, 6–7. https://nrcgt.uconn.edu/newsletters/winter934/

Feng, A. X. (2004). Constructing and implementing surveys. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 41–66). Prufrock Press.

Fink, A. (2003). *How to ask survey questions* (2nd ed.). Sage.

Florida Department of Education. (2017). *Florida school district gifted self-assessment tool, part I: Profile of the district*. www.fldoe.org/core/fileparse.php/7567/urlt/RGEGSF.pdf

Fowler, J. F. (2014). *Survey research methods* (5th ed.). Sage.

Focus Groups

Feng, A. X., & Brown, E. (2004). Using focus groups in gifted program evaluation. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 109–132). Prufrock Press.

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006a). *Advanced Placement and International Baccalaureate Programs: A "fit" for gifted learners* (Research Monograph 06222). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm06222.pdf

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006b). *AP student interview questions*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/AP-Student-Interview-Questions_Rev.pdf

Kitsantas, A., Bland, L., & Chirinos, D. S. (2017). Gifted students' perceptions of gifted programs: An inquiry into their academic and social-emotional functioning. *Journal for the Education of the Gifted*, *40*(3), 266–288. https://doi.org/10.1177%2F0162353217717033

Krueger, R. A., & Casey, M. A. (2001). Designing and conducting focus group interviews. In *Social analysis: Selected tools and techniques* (Paper Number 36, pp. 4–23). Social Development Family of the World Bank. http://documents1.worldbank.org/curated/en/568611468763498929/pdf/282790SDP136.pdf

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 school districts: A Clarion call for enhanced gifted program development. *Gifted Child Quarterly*, *50*(3), 339–341. https://doi.org/10.1177/001698620605000302

Interviews

Dai, D.Y., & Steenbergen-Hu, S. (2015). Special class for the gifted young: A 34-year experimentation with early college entrance programs in China. *Roeper Review*, *37*(1), 9–18. https://doi.org/10.1080/02783193.2014.975882

Dai, D. Y., Steenbergen-Hu, S., & Zhou, Y. (2015). Cope and grow: A grounded theory approach to early college entrants' lived experiences and changes in a STEM program. *Gifted Child Quarterly*, *59*(2), 75–90. https://doi.org/10.1177/0016986214568719

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006a). *Advanced Placement and International Baccalaureate Programs: A "fit" for gifted learners* (Research Monograph 06222). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm06222.pdf

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006c). *AP teacher interview questions*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/AP-Teacher-Interview-Questions_Rev.pdf

Hertzog, N. B. (2003). Impact of gifted programs from the students' perspectives. *Gifted Child Quarterly*, *47*(2), 131–143. https://doi.org/10.1177/001698620304700204

Rating Scales

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly*, *60*(2), 81–101. http://doi.org/10.1177/0016986216634438

Beasley, J. G., Briggs, C., & Pennington, L. (2017). Bridging the gap 10 years later: A tool and technique to analyze and evaluate advanced academic curricular units. *Gifted Child Today*, *40*(1), 48–58. https://doi.org/10.1177/1076217516675902

Colorado Department of Education. (2020). *Quality program assessment rubric*. www.cde.state.co.us/gt/giftededqualityprogramassessment

Cotabish, A., Shaunessy-Dedrick, E., Dailey, D., Keilty, B., Pratt, D., & Adams, C. (2020). *Self-assess your P-12 practice or program using the NAGC Gifted Programming Standards* (2nd ed.). National Association for Gifted Children.

Florida Department of Education. (2017). *Assessing the EP for students who are gifted.* www.fldoe.org/core/fileparse.php/7567/urlt/RGEGSF.pdf

Florida Department of Education. (2017). *Florida school district gifted program self-assessment tool, part II: Standards for gifted identification.* www.fldoe.org/core/fileparse.php/7567/urlt/RGEGSF.pdf

Gentry, M., & Gable, R. K. (n.d.). *My Class Activities: A survey instrument to assess students' perceptions of interest, challenge, choice, and enjoyment in their classrooms*. https://purduegeri.wixsite.com/instrument/my-class-activities

Gentry, M., & Gable, R. K. (2000). *My Class Activities*. Creative Learning Press. https://purduegeri.wixsite.com/instrument/my-class-activities

Gentry, M., & Gable, R. K. (2001). From the student's perspective—My Class Activities: An instrument for use in research and evaluation. *Journal for the Education of the Gifted*, *24*(4), 322–343. https://doi.org/10.1177%2F016235320102400403

Hunsaker, S. L., Frasier, M. M., Frank, E., Finley, V., & Klekotka, P. (1995). *Performance of economically disadvantaged students placed in gifted programs through the research-based assessment plan* (Research Monograph 95208). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95208.pdf

Pereira, N., Peters, S. J., & Gentry, M. (2010). My Class Activities instrument used in Saturday enrichment program evaluation. *Journal of Advanced Academics*, *21*(4), 568–593. https://doi.org/10.1177%2F1932202X1002100402

Purcell, J. H., Burns, D. E., Tomlinson, C. A., Imbeau, M. B., & Martin, J. L. (2002). Bridging the gap: A tool and technique to analyze and evaluate gifted education curricular units. *Gifted Child Quarterly*, *46*(4), 306–321. https://doi.org/10.1177/001698620204600407

Renzulli, J. S., & Westberg, K. L. (1991). *Scale for rating student's participation in the local gifted education program*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/Scales-for-Rating-Students-Participation-in-Gifted-Program_Rev.pdf

Spector, P. E. (1992). *Summated rating scale construction: An Introduction*. Sage.

Observation Protocols

Cassady, J. C., Speirs Neumeister, K. L., Adams, C. M., Cross, T. C., Dixon, F. A., & Pierce, R. L. (2004). The differentiated classroom observation scale, *Roeper Review*, *26*(3), 139–146. https://doi.org/10.1080/02783190409554259

Farah, Y. N., & Chandler, K. L. (2018). Structured observation instruments assessing instructional practices with gifted and talented students: A review of the literature. *Gifted Child Quarterly*, *62*(3), 276–288. https://doi.org/10.1177/0016986218758439

Hansen, J. B, & Feldhusen, J. F. (1994). Comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly*, *38*(3), 115–121. https://doi.org/10.1177/001698629403800304

Johnsen, S. K. (2016). Observation tool: Classroom observation of differentiation. In S. K. Johnsen, J. VanTassel-Baska, A. Robinson, A. Cotabish, D. Dailey, J. Jolly, J. Clarenbach, & C. Adams (Eds.), *Using the national gifted education standards for teacher preparation* (pp. 283–293). Prufrock Press.

Johnsen, S. K., Fearon-Drake, D., & Wisely, L. W. (2020). A formative evaluation of differentiation practices in elementary cluster classrooms, *Roeper Review*, *42*(3), 206–218. https://doi.org/10.1080/02783193.2020.1765921

Johnsen, S. K., Haensley, P. S., Ryser, G. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*(1), 45–63. https://doi.org/10.1177/001698620204600105

Johnsen, S. K., & Thompson, L. W. (2014). *An observation report of fourth grade cluster classrooms in Waco ISD: Relationships to student achievement*. Baylor University.

Johnsen, S. K., Thompson, L., Scalzo, R., & Farah, Y. (2013). *A follow-up observation report of elementary cluster classrooms in Waco ISD*. Baylor University.

Pereira, N., Tay, J., Desmet, O., Maeda, Y., & Gentry, M. (2021). Validity evidence for the revised Classroom Practices Survey: An instrument to measure teachers' differentiation practices. *Journal for the Education of the Gifted*, *44*(1), 31–55. https://doi.org/10.1177/0162353220978304

Pereira, N., Tay, J., Maeda, Y., & Gentry, M. (2019). Differentiation as measured by the Classroom Practices Survey: A validity study updating the original instrument. *Learning Environments Research*, *22*, 443–460. https://doi.org/10.1007/s10984-019-09284-z

Peters, S. J., & Gates, J. C. (2010). The Teacher Observation Form: Revisions and updates. *Gifted Child Quarterly*, *54*(3), 179–188. https://doi.org/10.1177/0016986210369258

Tomlinson, C. A., Callahan, C. M., Moon, T. R., Tomchin, E. M., Landrum, M., Imbeau, M., Hunsaker, S. L., & Eiss, N. (1995). *Preservice teacher preparation in meeting the needs of gifted and other academically diverse students*. (Research Monograph 95134). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95134.pdf

VanTassel-Baska, J. (2004). Assessing classroom practice: The use of a structured observation form. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 87–107). Prufrock Press.

VanTassel-Baska, J., Avery, L., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). *The William and Mary Classroom Observation Scales—Revised*. Center for Gifted Education. https://education.wm.edu/centers/cfge/_documents/research/athena/cosrform.pdf

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2005). *Classroom Observation Scales–Revised: User's manual*. Center for Gifted Education. https://education.wm.edu/centers/cfge/_documents/resources/cosrmanualrevised.pdf

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2006). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review*, *29*(2), 84–92. https://doi.org/10.1080/02783190709554391

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). *An observational study of instructional and curricular practices used with gifted and talented students in regular classrooms* (Research Monograph 93194). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm93104.pdf

Document Review

Ford, D. Y. (2009). *Multicultural gifted education: Providing mirrors for ALL students*. https://12d9e5b0-b4ba-b916-f1ca-20b09d90331f.filesusr.com/ugd/55c01c_2614888efbb56329befece0fbdaa2336.pdf

Maryland State Department of Education and The Maryland State Advisory Council. (2021). *Maryland's 2021 EGATE application*. http://marylandpublicschools.org/programs/Documents/Gifted-Talented/2021EGATEApplicationDirections.pdf

VanTassel-Baska, J., & Baska, A. (2019). Appendix F: Curriculum review checklist. *Curriculum planning and instructional design for gifted learners*. (3rd ed., pp. 281–291). Prufrock.

Virginia Department of Education. (2013). *Technical review of the local plan for the education of the gifted*. www.doe.virginia.gov/instruction/gifted_ed/technical_review_of_local_gifted_plan.pdf

Templates

Delcourt, M. A. B. (1994). *Program profile form*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Program-Profile-Form_Rev.pdf

Eckert, R. D., & Robins, J. H. (Eds.). (2017). *Designing services and programs for high-ability learners: A guidebook for gifted education* (2nd ed.). Corwin. https://doi.org/10.4135/9781506343440

Pflugerville Independent School District. (n.d.) *Gifted and talented program review and gap analysis chart*. www.pfisd.net/cms/lib/TX01001527/Centricity/Domain/1294/Program%20Review%20and%20Gap%20Analysis%20Chart14.15.pdf

Appendix D

Specific Examples of Evaluation Tools

Name	Type of Tool	How the Tool is Used for Evaluation	Empirical Sources (see References)	Psychometric Qualities	Use with Gifted Students	Use with Diverse Students	NAGC Standard Alignment	Online Access
Self-Study Checklist	Rating Scale	Evaluate program using the NAGC Programming Standards.	Cotabish et al. (2017, 2020)	Content validity: aligned to NAGC standards and theory, research, and practice	Yes	Yes	2019	Example: www.nagc.org/sites/default/files/WebinarPowerPoints/SelfAssess%20P12%20Programming.pdf
Gap Analysis Chart	Template	To "determine the steps needed to move from a current state of implementation or development to a future desired state" (Cotabish et al., 2017, p. 7)	Cotabish et al. (2017); Pflugerville Independent School District (n.d.)	Content validity: should be aligned to NAGC standards and theory, research, and practice	Yes	Yes	2010 No (Texas standards)	www.pfisd.net/cms/lib/TX01001527/Centricity/Domain/1294/Program%20Review%20and%20Gap%20Analysis%20Chart14.15.pdf
Classroom Instructional Practices Scale	Observation Protocol	To rate a teacher's classroom practices using data that have been gathered from classroom observation of a learning task, interviews, questioning, and room arrangement (Johnsen, 2016, Johnsen et al., 2020).	Johnsen (2016); Johnsen et al. (2002, 2013, 2020); Johnsen and Thompson (2014); Ryser and Johnsen (1996a, 1996b)	Inter-rater reliability: .92 to .95 Content validity: observation Internal consistency: 0.76 to 0.795 Evaluator agreement: 93% to 95%	Yes	Yes	2019	The classroom observation instrument, interview forms, and CIPS can be obtained from Susan Johnsen (email: susan_johnsen@baylor.edu).

Classroom Observation Scales—Revised	Observation Protocol	To assess teacher performance in response to gifted learners for a program evaluation (Avery et al., 1997).	Avery et al. (1997); VanTassel-Baska (2004); VanTassel-Baska et al. (2003, 2005, 2006, 2020)	Interrater reliability: .87 to .89 Internal consistency: .91 to .93 Content validity: .98	Yes	Yes	2010	Copies of both the manual and protocol are available from the Center for Gifted Education or the primary author (jlvant@wm.edu) https://education.wm.edu/centers/cfge/_documents/research/athena/cosrform.pdf https://education.wm.edu/centers/cfge/_documents/resources/cosrmanualrevised.pdf
Program Satisfaction Surveys	Survey	To design programs, compare several programs, and evaluate key components of an established program.	Delcourt and McIntire (1993); Delcourt and Evans (1994c)	Content validity: expert review	Yes	Yes	No	**Parent** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Parent-Satisfaction-Survey_Rev.pdf **Student** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Student-Satisfaction-Survey_Rev.pdf **Teacher** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Teacher-Satisfaction-Survey_Rev.pdf **Administrator** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Administrator-Satisfaction-Survey_Rev.pdf

(*continued*)

Name	Type of Tool	How the Tool is Used for Evaluation	Empirical Sources (see References)	Psychometric Qualities	Use with Gifted Students	Use with Diverse Students	NAGC Standard Alignment	Online Access
AP and IB Student Interview Questions	Focus Group	To understand teachers' conceptualization and implementation of curriculum and instruction for gifted learners in AP/IB classes and how students enrolled in the classes perceive and evaluate their learning experiences.	Hertberg-Davis et al. (2006a)	Content validity: aligned with best practices in qualitative research and gifted education	Yes	Yes	No	**AP** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/AP-Student-Interview-Questions_Rev.pdf **IB** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/IB-Student-Interview-Questions_Rev.pdf
AP and IB Teacher Interview Questions	Interview						No	**AP** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/AP-Teacher-Interview-Questions_Rev.pdf **IB** https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/04/IB-Teacher-Interview-Questions_Rev.pdf

References

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly*, *60*(2), 81–101. http://doi.org/10.1177/0016986216634438

American Psychological Association. (2014). *Assessing and evaluating teacher preparation programs*. Center for Psychology in Schools and Education. www.apa.org/ed/schools/teaching-learning/teacher-preparation-programs.pdf

Avery, L. D., VanTassel-Baska, J., & O'Neill, B. (1997). Making evaluation work: One school district's experience. *Gifted Child Quarterly*, *41*(4), 124–131. https://doi.org/10.1177/001698629704100402

Baldwin, A. Y. (2002). Culturally diverse students who are gifted. *Exceptionality*, *10*(2), 139–147. https://doi.org/10.1207/S15327035EX1002_7

Beasley, J. G., Briggs, C., & Pennington, L. (2017). Bridging the gap 10 years later: A tool and technique to analyze and evaluate advanced academic curricular units. *Gifted Child Today*, *40*(1), 48–58. https://doi.org/10.1177/1076217516675902

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Callahan, C. M. (2009). Evaluation for decision-making: The practitioner's guide to program evaluation. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted* (pp. 119–142). Prufrock.

Callahan, C. M. (2017). Developing a plan for evaluating services provided to gifted students. In R. D. Eckert & J. H. Robins (Eds.), *Designing services and programs for high-ability learners: A guidebook for gifted education* (2nd ed., pp. 225–238). Corwin.

Callahan, C. M. (2020). Gifted programming standards: Understanding their role in program design and evaluation. In J. H. Robins, J. L. Jolly, F. A. Karnes, & S. M. Bean (Eds.), *Methods & materials for teaching the gifted* (5th ed., pp. 127–144). Prufrock.

Callahan, C. M., Moon, T. R., & Oh, S. (2017). Describing the status of programs for the gifted: A call to action. *Journal for the Education of the Gifted*, *40*(1), 20–49. https:// doi.org/10.1177/0162353216686215

Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instrument and evaluation designs used in gifted programs* (Research Monograph 95132). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95132.pdf

Cassady, J. C., Speirs Neumeister, K. L., Adams, C. M., Cross, T. C., Dixon, F. A., & Pierce, R. L. (2004). The differentiated classroom observation scale. *Roeper Review*, *26*(3), 139–146. https://doi.org/10.1080/02783190409554259

Cotabish, A., Dailey, D., & Jackson, N. (2017). Aligning gifted programming and services with national and state standards. In R. D. Eckert & J. H. Robins (Eds.), *Designing services and programs for high-ability learners: A guidebook for gifted education* (2nd ed., pp. 1–17). Corwin.

Cotabish, A., Shaunessy-Dedrick, E., Dailey, D., Keilty, B., Pratt, D., & Adams, C. (2020). *Self-assess your P-12 practice or program using the NAGC Gifted Programming Standards* (2nd ed.). National Association for Gifted Children.

Council for Exceptional Children, The Association for the Gifted. (2020). *A call to action: Supporting equity, diversity, and access for gifted students*. http://cectag.com/wp-content/uploads/2021/06/Call-to-Action-2020-with-All-Endorsements.pdf

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Davis, J. L. (2019). Reframing professional learning to meet the needs of teachers working with culturally diverse gifted learners. In A. M. Novak & C. L. Weber (Eds.), *Best practices in professional learning and teacher preparation: Special topics for gifted professional development* (Vol. 2, pp. 51–70). Prufrock.

Delcourt, M. A. B., & Evans, K. (1994c). *Qualitative extension of the learning outcomes study* (Research Monograph 94110). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm94110.pdf

Delcourt, M. A. B., & McIntire, J. A. (1993). An investigation of student learning outcomes: Results of a program satisfaction survey. *Newsletter of The National Research Center on the Gifted and Talented, Winter*, 6-7. https://nrcgt.uconn.edu/newsletters/winter934/

Delcourt, M. A. B., & Evans, K. (1994b). *Parent satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Parent-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & Evans, K. (1994d). *Student satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Student-Satisfaction-Survey_Rev.pdf

Delcourt, M. A. B., & Evans, K. (1994e). *Teacher satisfaction survey*. National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2021/05/Teacher-Satisfaction-Survey_Rev.pdf

Farah, Y. N., & Chandler, K. L. (2018). Structured observation instruments assessing instructional practices with gifted and talented students: A review of the literature. *Gifted Child Quarterly*, *62*(3), 276–288. https://doi.org/10.1177/0016986218758439

Feng, A. X. (2004). Constructing and implementing surveys. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 41–66). Prufrock.

Feng, A. X., & Brown, E. (2004). Using focus groups in gifted program evaluation. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 109–132). Prufrock.

Fink, A. (2003). *How to ask survey questions* (2nd ed.). Sage.

Ford, D. Y. (2009). *Multicultural gifted education: Providing mirrors for ALL students*. https://12d9e5b0-b4ba-b916-f1ca-20b09d90331f.filesusr.com/ugd/55c01c_2614888efbb56329befece0fbdaa2336.pdf

Ford, D. Y., Dickson, K. T., Davis, J. L., Scott, M. T., & Grantham, T. C. (2018). A culturally responsive equity-based bill of rights for gifted students of color. *Gifted Child Today*, *41*(3), 125–129. https://doi.org/10.1177/1076217518769698

Fowler, J. F. (2014). *Survey research methods* (5th ed.). Sage.

Gentry, M., & Gable, R. K. (2001). From the student's perspective—My Class Activities: An instrument for use in research and evaluation. *Journal for the Education of the Gifted*, *24*(4), 322–343. https://doi.org/10.1177%2F016235320102400403

Grissom, J. A., & Redding, C. (2017). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open 2*(1), 1–25. https://doi.org/10.1177/2332858415622175

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3). https://doi.org/10.1080/135406002100000512

Hansen, J. B, & Feldhusen, J. F. (1994). Comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly*, *38*(3), 115–121. https://doi.org/10.1177/001698629403800304

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006a). *Advanced Placement and International Baccalaureate Programs: A "fit" for gifted learners* (Research Monograph 06222). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm06222.pdf

Johnsen, S. K. (2016). Observation tool: Classroom observation of differentiation. In S. K. Johnsen, J. VanTassel-Baska, A. Robinson, A. Cotabish, D. Dailey, J. Jolly, J. Clarenbach, & C. Adams (Eds.), *Using the national gifted education standards for teacher preparation* (pp. 283–293). Prufrock.

Johnsen, S. K., Fearon-Drake, D., & Wisely, L. W. (2020). A formative evaluation of differentiation practices in elementary cluster classrooms, *Roeper Review*, *42*(3), 206–218. https://doi.org/10.1080/02783193.2020.1765921

Johnsen, S. K., Haensley, P. S., Ryser, G. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*(1), 45–63. https://doi.org/10.1177/001698620204600105

Johnsen, S. K., & Thompson, L. W. (2014). *An observation report of fourth grade cluster classrooms in Waco ISD: Relationships to student achievement*. Baylor University.

Johnsen, S. K., Thompson, L., Scalzo, R., & Farah, Y. (2013). *A follow-up observation report of elementary cluster classrooms in Waco ISD.* Baylor University.

Krueger, R. A., & Casey, M. A. (2001). Designing and conducting focus group interviews. In *Social analysis: Selected tools and techniques* (Paper Number 36, pp. 4–23). Social Development Family of the World Bank. http://documents1.worldbank.org/curated/en/568611468763498929/pdf/282790SDP136.pdf

Lakin, J. M., & Rambo-Hernandez, K. (2019). Fidelity of implementation: Understanding why and when programs work. *Gifted Child Today*, *42*(4), 205–214. https:/doi.org/10.1177/1076217519862327

Maryland State Department of Education and The Maryland State Advisory Council. (2021). *Maryland's 2021 EGATE application.* http://marylandpublicschools.org/programs/Documents/Gifted-Talented/2021EGATEApplicationDirections.pdf

Mun, R. U., Ezzani, M. D., & Lee, L. E. (2020). Culturally relevant leadership in gifted education: A systematic literature review. *Journal for the Education of the Gifted*, *43*(2), 108–142. https://doi.org/10.1177%2F0162353220912009

National Association for Gifted Children. (2010). *The NAGC pre-K to grade 12 gifted programming standards.* www.nagc.org/sites/default/files/standards/K-12%20programming%20standards.pdf

National Association for Gifted Children. (2019). *The NAGC pre-K to grade 12 gifted programming standards.* www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20Standards%281%29.pdf

National Association for Gifted Children, & Council for Exceptional Children, The Association for the Gifted. (2013). *Teacher preparation standards in gifted and talented education.* www.nagc.org/sites/default/files/standards/NAGC-%20CEC%20CAEP%20standards%20(2013%20final).pdf

National Research Center on the Gifted and Talented. (n.d.-a). *NRC/GT Instrument Bank Instruments.* University of Connecticut. https://nrcgt.uconn.edu/research-based_resources/ibcd05/

National Research Center on the Gifted and Talented. (n.d.-b). *NRC/GT Instrument Bank Instruments II.* University of Connecticut. https://nrcgt.uconn.edu/research-based_resources/ibcd09/

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation* (4th ed.). Wiley.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102%2F0034654307313793

Patton, M. Q. (2015). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). Sage.

Pereira, N., Peters, S. J., & Gentry, M. (2010). My Class Activities instrument used in Saturday enrichment program evaluation. *Journal of Advanced Academics*, *21*(4), 568–593. https://doi.org/10.1177%2F1932202X1002100402

Pereira, N., Tay, J., Desmet, O., Maeda, Y., & Gentry, M. (2021). Validity evidence for the revised Classroom Practices Survey: An instrument to measure teachers' differentiation practices. *Journal for the Education of the Gifted*, *44*(1), 31–55. https://doi.org/10.1177/0162353220978304

Pereira, N., Tay, J., Maeda, Y., & Gentry, M. (2019). Differentiation as measured by the Classroom Practices Survey: A validity study updating the original instrument. *Learning Environments Research*, *22*, 443–460. https://doi.org/10.1007/s10984-019-09284-z

Peters, S. J., & Gates, J. C. (2010). The Teacher Observation Form: Revisions and updates. *Gifted Child Quarterly*, *54*(3), 179–188. https://doi.org/10.1177/0016986210369258

Peters, S. J. & McBee, M. (2019, November). *Implementing local norms: A step-by-step guide.* Session presented at the annual meeting of the National Association for Gifted Children (NAGC). Albuquerque, NM. https://osf.io/jbasv/

Pflugerville Independent School District. (n.d.) *Gifted and talented program review and gap analysis chart.* https://www.pfisd.net/cms/lib/TX01001527/Centricity/Domain/1294/Program%20Review%20and%20Gap%20Analysis%20Chart14.15.pdf

Purcell, J. H., Burns, D. E., Tomlinson, C. A., Imbeau, M. B., & Martin, J. L. (2002). Bridging the gap: A tool and technique to analyze and evaluate gifted education curricular units. *Gifted Child Quarterly*, *46*(4), 306–321. https://doi.org/10.1177/001698620204600407

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the States in Gifted Education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. www.nagc.org/2018-2019-state-states-gifted-education

Robbins, J. I. (2019). Gifted program evaluation: Catalyst for change or keeper of the culture. *Gifted Child Today*, *42*(4), 196–204. https://doi.org/10.1177%2F1076217519862326

Ryser, G. R. (2018a). Fairness in testing and nonbiased assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 59–72). Prufrock.

Ryser, G. R. (2018b). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide* (3rd ed., pp. 33–58). Prufrock.

Ryser, G. R., & Johnsen, S. K. (1996a). An overview of effective practices with gifted students in general education settings. *Journal for the Education of the Gifted*, *19*(4), 379–404. https://doi.org/10.1177/016235329601900402

Ryser, G., & Johnsen, S. K. (1996b). Toward more research on effective practices with gifted students in general education settings. *Journal for the Education of the Gifted*, *19*(4), 481–496. https://doi.org/10.1177/016235329601900406

Spector, P. E. (1992). *Summated rating scale construction: An Introduction*. Sage.

Speirs Neumeister, K., & Burney, V. H. (2019). *Gifted program evaluation: A handbook for administrators and coordinators* (2nd ed.). Prufrock.

Tomlinson, C. A., Callahan, C. M., Moon, T. R., Tomchin, E. M., Landrum, M., Imbeau, M., Hunsaker, S. L., & Eiss, N. (1995). *Preservice teacher preparation in meeting the needs of gifted and other academically diverse students*. (Research Monograph 95134). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95134.pdf

VanTassel-Baska, J. (2004). Assessing classroom practice: The use of a structured observation form. In J. VanTassel-Baska & A. X. Feng (Eds.), *Designing and utilizing evaluation for gifted program improvement* (pp. 87–107). Prufrock Press.

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 school districts: A clarion call for enhanced gifted program development. *Gifted Child Quarterly*, *50*(3), 339–341. https://doi.org/10.1177/001698620605000302

VanTassel-Baska, J., Avery, L., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). *The William and Mary Classroom Observation Scales—Revised*. Center for Gifted Education. https://education.wm.edu/centers/cfge/_documents/research/athena/cosrform.pdf

VanTassel-Baska, J., & Baska, A. (2019). Appendix F: Curriculum review checklist. *Curriculum planning and instructional design for gifted learners* (3rd ed., pp. 281–291). Prufrock.

VanTassel-Baska, J., Hubbard, G. F., & Robbins, J. I. (2020). Differentiation of instruction for gifted learners: Collated evaluative studies of teacher classroom practices. *Roeper Review*, *42*(3), 153–164. https://doi.org/10.1080/02783193.2020.1765919

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2005). *Classroom Observation Scales–Revised: User's manual*. Center for Gifted Education. https://education.wm.edu/centers/cfge/_documents/resources/cosrmanualrevised.pdf

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2006). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review*, *29*(2), 84–92. https://doi.org/10.1080/02783190709554391

Virginia Department of Education. (2013). *Technical review of the local plan for the education of the gifted*. www.doe.virginia.gov/instruction/gifted_ed/technical_review_of_local_gifted_plan.pdf

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). *An observational study of instructional and curricular practices used with gifted and talented students in regular classrooms* (Research Monograph 93194). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm93104.pdf

CHAPTER

16

Case Studies of District Models for Success

Carolyn M. Callahan

Case Studies of District Models for Success

The success of evaluation efforts to assess the processes and products of gifted programming components rests on four critical elements: (a) the degree to which the evaluation design and implementation addresses important (to stakeholders) questions, concerns and issues, (b) the degree to which the data collected are valid and reliable indicators to answer those questions, (c) the degree to which a school district actually implements recommended changes as they are supported by the data and conclusions of the evaluator(s) and (d) the degree to which those changes result in improvements to program operations and outcomes.

While program evaluation recommendations should be based on many sources and types of data, two of the most crucial sources in making sound recommendations are student data and teacher data. But not just any data will do. The student and teacher data collected in the process of program evaluation must be clearly related to the evaluation issues, questions, and concerns that have been targeted for investigation in the evaluation. That is, data collection instruments must yield valid information for answering the question, and the data must be reliable. Finally, the data must have "face validity" for the audiences who will receive the findings and recommendations based on those findings. For stakeholders to see a compelling argument for either maintaining the status quo or for making changes, the findings must offer evidence *that is convincing to them*.

Types of Data

As noted above, examination of the adequacy of resources, processes and products of a gifted program will likely rest on the examination of an array of data ranging from qualitative analysis of program documents to qualitative and quantitative data about or from students, qualitative or quantitative data about or from teachers, classroom observations, and qualitative and quantitative data from parents, administrators, counselors, school board members and other stakeholders in the success of the program. (See pp. 272–73 for criteria used in selecting and interpreting instruments of value in this process.)

DOI: 10.4324/9781003285991-24

Further, data collected are often categorized as formative or summative. That is, some data (formative) are collected to provide information on factors that are likely to positively or negatively impact outcomes (e.g., data may be collected to ascertain whether teachers perceive they have adequate understanding of the curriculum they are to teach, based on the assumption that instruction will not produce the desired outcomes unless they do). Or we might assess the whether the curricular and instructional activities in teachers' classrooms reflect the key elements of the professional development they have been provided. If not, then adjustments to professional development are in order because we cannot expect that students' learning will reflect the goals of the curriculum if it is not implemented. Summative data are used to judge the outcomes of providing appropriate resources and carrying out the designated operations/activities of a program. For example, does student learning reflect growth in critical thinking? Do students learn more advanced concepts in mathematics than they would have without participation in the program?

Student Data

Student data can inform the identification process, the curriculum and instructional practice, and the outcomes of the curricular and instructional implementation. Student demographic data might be used to evaluate the effectiveness, efficiency, and equity of identification procedures. Student outcome data used to evaluate the outcomes of instruction can be scores on standardized or locally constructed objective or open-ended tests, ratings on performance or product assessments, or observations of classrooms. Students may also be the source of data collected via surveys/questionnaires and/or interviews (individual or focus group) across a variety of dimensions such as satisfaction with the curriculum and/or instruction, classroom climate, and the challenge of the curriculum they are offered, among other factors.

Teacher Data

Teachers are also a valuable source of data in assessing program effectiveness. The data they provide can take the form of direct input on such factors as their perceptions of the effectiveness of professional learning activities, the adequacy of resources for instruction, student behaviors including student performance, the administrative support available, or program operations in general. Indirectly, teachers may provide data through observation of their classrooms with a focus on curricular differentiation, instructional strategies, classroom climate, etc.

Sometimes we must consider both student and teacher data to come to a full understanding of what works and what must be altered to maximize the effectiveness of the process and procedures in programming. Student and teacher data used in conjunction with one another may either provide confirmation of conclusions (one source affirming the other—referred to as one aspect of triangulation) or one data set may provide information that contributes to understanding the results presented in the other data set. Consider the case of Anderson County Programs for the Gifted and Talented. This first case study and those that follow are representative of particular aspects of gifted programming. In doing a full program evaluation the evaluator would engage in a more comprehensive examination of all aspects of the program rather than just a focus on one aspect of the program.

Case 1

Anderson City Schools had an identification process which incorporated both their long-standing traditional identification process as Path 1 and a second alternative process as Path 2. The traditional process relied on an evaluation of second grade students' scores based on national norms derived from a traditional test of cognitive processes (with a verbal, quantitative and spatial reasoning component) and student scores from a rating scale on which teachers rate a set of student behaviors considered indicative of giftedness. Only students who were referred by their teachers or their parents were administered the cognitive assessment, and teachers only completed the rating scales on the nominated students. Scores on the assessments were examined by a panel; only those students meeting pre-set criteria on both assessments were offered services in the gifted program. In the newly implemented identification process, Path 2, *all* second-grade students were assessed on a test of cognitive abilities, and teachers completed the rating scales on *all* second-grade students (universal screening with both instruments). Local and national norms on the cognitive ability test and classroom and district norms on the teacher rating scales were presented to a review committee. This panel, comprised of administrators and teachers, reviewed the profiles to determine who would be offered gifted services based on their professional judgment that these students would benefit from the gifted program offered.

The school district administrators wished to evaluate this two-path identification process. The school district staff and the evaluator determined that there should be a focus on four evaluation questions:

1. Do school staff involved in the identification process understand and feel comfortable with the implementation of the revised identification process?
2. Does the incorporation of Path 2 in the identification process lead to representation in the gifted program that more closely reflects the demographics of the general population in the school district across race/ethnicity, socio-economic status, EL status, status as a student identified with a disability?
3. Do students selected for the program accept and continue enrollment in the program across all categories named above?
4. Does the process yield a pool of students who are successful in the curriculum offered in the program for gifted and talented students? The curriculum offered to gifted students was characterized in program documents as an enrichment program with an emphasis on developing critical and creative thinking and problem-solving across disciplines.

Data

To address the evaluation questions the evaluator collected both student data and teacher data) including:

- data from interviews with the administrators and teachers on the selection committees
- the number of students identified in each of the underrepresentation categories (race/ethnicity, socio-economic status, EL status, status as a student identified with

a disability) identified in evaluation question 2 and the number of students in each category in the student population overall,

- the retention rate over 2 years from time of first enrollment across all groups of students,
- data gathered via surveys of students' perceptions of the classroom climate, the curriculum, and the instructional strategies in the gifted program,
- students' scores on the Diet Cola Test (Fowler, 1990), a measure of problem solving in science, administered at the beginning and end of third grade to identified students.
- student scores based on a rubric designed to assess critical thinking, creativity, and advanced content used to rate products produced during the culminating activity in an inter-disciplinary problem-based learning unit (the rubric was a modified version of the Student Product Assessment Form (SPAF; Renzulli & Reis, https://gifted.uconn.edu/wp-content/uploads/sites/961/2015/02/spaf.pdf).

Teacher data included responses to a survey given to all teachers in which they were asked to indicate their perceptions of the readiness of the identified students to succeed in the curriculum offered, their knowledge of strategies for scaffolding their curriculum, and their perceptions of the effectiveness of professional learning sessions. Small group focus interviews provided additional data on teacher perceptions.

Results

The data related to student representation revealed that the traditional strategy (Path 1) did not reflect proportional data relative to the district's total school demographics. These data were not surprising, given the prior record of the outcomes of that identification process which had been the impetus for adding Path 2. The result of implementing Path 2 procedures was a pool of additional students whose demographics more closely reflected the overall demographics of the district but were not an exact match. The proportion of identified African American and Latinx (not EL) students and students who qualified for free and reduced meals (FARMS) increased significantly, but the groups who were still under-represented and for whom there was only a minimal increase in proportionality of students identified were students who were ELs and students with disabilities. Those individuals who participated in the identification process indicated that they understood and felt comfortable with the process.

However, the data on retention indicated that the students identified by Path 1 had a negligible rate of leaving the program, while among Path 2 students the percentage of those leaving the program was nearly 10%. No statistically significant differences were detected between students who continued in the program and those who chose to leave the program on scores earned on the cognitive ability assessment nor on the teacher ratings of students.

Analysis of the student survey responses indicated high levels of satisfaction, sense of belonging, engagement, and enjoyment of the learning activities offered in the gifted program among Path 1 students, but relatively low levels of satisfaction, sense of belonging, engagement, and enjoyment among Path 2 students. Students in the Path 1 and Path 2 groups of students who remained in the program earned statistically similar scores on the Diet Cola Test at the conclusion of year 1. Scores on the Diet Cola Test were higher on the post-test than on the pre-test. Students' ratings on the SPAF did not differ significantly between those in Path 1 and those in Path 2.

The teachers' responses to surveys and interviews indicated that many of the teachers did not feel the Path 2 students as a group (with some noted exceptions) had the requisite skills and knowledge to succeed in the program and did not feel prepared to provide appropriate scaffolding to modify their lessons to accommodate the level of preparation of students or to make culturally relevant modifications to the curriculum or to instructional practice. The professional learning sessions were not judged to be useful in providing clear and applicable information on scaffolding, culturally relevant curriculum, or instructional practice. Further, several teachers recommended that the identification process revert to the more traditional process.

The findings about retention and student level of satisfaction obviously provided important indicators of areas of concern. First, the differential between the number of students in Path 1 and the number of students in Path 2 who chose not to continue in the program is high, and the 10% "dropout" rate (Path 2 students) is exceptional for a gifted program. Combining the retention issues with the lower levels of satisfaction, enjoyment, engagement, and sense of belonging suggests that the program and classroom activities are less fitting for the Path 2 students. The results of the teacher surveys and interviews offered potential explanations for those results. Teachers in the gifted program had apparently not been prepared for the inclusion of a more diverse population of students in the gifted program. The teachers seemed to lack either the skill and/or the will to accommodate increased diversity within the instructional offerings in their classrooms.

Recommendations

Four major recommendations emerged from these data.

1. The district should apply the process used for Plan 2 students to all students in the district. That process reflects recommended practices of universal screening and application of local norms to the scores from that screening, as described in current literature on the identification of gifted students (e.g., Peters & Engerrand, 2016; Peters et al., 2020), and examination of the data indicated nearly all students identified by the first process would likely be included by applying the second process to all students.
2. The district should consider revisiting the identification process to incorporate the recommendations experts offer of strategies to identify EL learners and students with disabilities (e.g., Lohman et al., 2008; Mun et al., 2020; Rizza & Morrison, 2007). The professional development program for the teachers should target greater attention to informing teachers of the appropriateness of the program for a more diverse population of students, with an emphasis on how opportunity to learn affects performance and how gifts and talents manifest in different ways in differing groups of students.
3. Teachers should be provided with targeted training on diversity in gifted students, culturally responsive teaching, scaffolding of curriculum (clearly as not "teaching down" or "watering down" the curriculum). They should be provided with resources both in terms of support and mentoring from administrators on scaffolding and

physical resources (books, curriculum, instructional materials) that will assist them in engaging learners from diverse populations.

4. Finally, the administrative staff should monitor the degree to which teachers make changes in curriculum and instructional practice that reflect the training and resources provided.

Results of Implementation of the Recommendations

The results of the implementation of Plan 2 with all students was the identification of a diverse overall population of students, approaching reflection of the demographics of the district across the dimensions of race/ethnicity, socio-economic statuses, EL learners, and twice-exceptional students. The retention rate of students fell overall to less than 1% of the identified students with no differences across race, socio-economic status, EL or status as students with disabilities. The students reported greater overall enjoyment and engagement in the gifted program.

One teacher chose to leave the position as teacher of the gifted when the new plan was implemented districtwide. The opening was filled with a bilingual teacher with background in EL instruction. The professional development sessions received high ratings in the areas that had been lacking in the prior assessment (scaffolding, culturally relevant curriculum, and instruction), teachers were more positive about the students identified, and they were appreciative of the physical resources offered. Unfortunately, the mentoring in scaffolding and evaluation of implementation of the expected changes in classrooms did not occur suggesting, perhaps, that the recommendation was not needed if the professional learning sessions were of sufficiently high quality.

Results on the Diet Cola Test and the SPAF were replicated across all demographic subgroups of students.

Case 2

At Schyler High School the teachers and administrators successfully expanded enrollment in the Advanced Placement program. As a result of an active recruitment program involving teachers and counselors and data from the PSAT, more students enrolled in AP courses last year than ever before, but the teachers and the students were disappointed in the low scores earned by many of the students who were heavily recruited to participate in the classes (only 20% earned a score of 3 or higher, which is generally considered as necessary to earn college credit). The teachers in the AP classes created a new summer program, the AP Passage Program, in conjunction with the local university teacher education program, to provide students with the prerequisite skills they felt students were lacking as they began their AP studies. The program involved a week of professional development sessions offered by the university staff to the AP teachers, which focused on content knowledge and strategies for teaching skills in reading, writing, and study skills. They also developed activities designed to enhance growth mindset in high school students. The AP teachers were also involved in developing scaffolding activities in areas where they had observed students struggling in the particular classes they taught. A new group of students was recruited, and the high school counselors invited all those identified as more "at risk" to attend a two-week program at the university with classes taught by the AP teachers they would have the next academic year and who had attended the professional development program. An evaluator was asked to collect data to assess the success of the intervention designed to improve student performance. The evaluation questions posed were:

1. To what extent did students participate in the summer program? Was it accessible and accessed by the students identified by counselors as candidates for the program?
2. Do the teachers perceive the sessions offered by university staff increased their ability to provide appropriate background knowledge, skills in reading and writing, and study skills to the students attending the summer program?
3. Did students perceive the summer program provided them with content knowledge and skills that were new and that contributed to their ability to succeed in the Advanced Placement courses in which they enrolled?
4. Did students' growth mindset improve as a result of participation in the summer program and teachers' efforts to enhance student growth mindset?
5. Did student performance in class activities and assessments improve as a result of the summer program and teacher adjustments to the curriculum and instructional practice?
6. Were the scores of students who participated in the summer program on the Advanced Placement exams higher than the scores of recruited students from the prior year (who had not opportunity to participate in a summer program) and were their scores higher than those of students who had been recruited but who chose not to participate in the summer program?

Data

To assess the AP Passage program, the evaluators collected teacher survey data following the professional development sessions and student data during and following the summer program, during the following academic year, and at the end of the academic year. Student data gathered during and following the summer program included:

- Attendance at the AP Passage Program
- Responses to surveys, focusing on student perceptions of the AP Passage experience

Student data collected during the academic year comprised:

- Data from interviews with students who attended the AP Passage program and students who had been invited, but who had not attended the AP Passage program about their perceptions of the AP classes during the academic year and their sense of preparedness for the classes.
- Scores on Implicit Theories of Intelligence Scale (Dweck, 2000) administered in September and in May as a measure of growth mindset.

Student data collected at the end of the academic year were:

- Scores on AP exams
- Grades in AP classes.

Results

Teachers who attended the summer training session reported that the sessions provided were useful in general but indicated that the sessions did not provide sufficient focus on strategies for providing scaffolding in the specific course(s) they were teaching or for motivating students to commit the time necessary to be successful in the AP class. They also expressed doubt about their ability to provide scaffolding while keeping all students at the same level of achievement and proceeding at the same pace of learning as other students in the class.

Approximately 75% of the invited students attended the AP Passage Program (attendance was *not* required for enrollment in AP classes). Students who chose not to attend the summer program provided several reasons for their decisions. First and foremost was financial need to work during the summer and inability to find positions where they could take a two-week hiatus from their jobs; second was an expressed confidence that they did not "need" additional preparation for the AP courses in which they enrolled; and third was an expressed reluctance to be "singled out" as an at-risk learner.

Of those students who attended the AP Passage summer program, nearly all reported on the end of program survey that the sessions were challenging, presented new information, and provided interest in the discipline of the AP class. During interviews with students who had and had not attended the summer program, three important themes emerged. The first was that the students who had attended the program expressed greater confidence in their ability to succeed. The second theme could be expressed as these same students were "embarrassed" by being singled out by teachers during class with comments such as "Don't you remember when we talked about this during the summer?" or "Adam can explain this because we talked about this during the summer orientation to the class." Finally, the students expressed conflicting feelings. While students who attended the program felt they learned useful content and skills, they also expressed concerns that could be summarized as feeling that if they were selected as needing this program that there was an accompanying "seed of doubt planted about their ability and/or preparedness for AP classes."

Scores did not change for any group of students on the growth mindset scale, but the pretest indicated most students expressed a relatively high growth mindset already.

AP scores were marginally higher for the highly recruited students than the prior year, with students who attended the AP Passage program scoring significantly higher than those who did not attend. However, only 35% of students in the highly recruited group who attended the summer program earned scores of 3 or higher (20% of recruited students from the prior year had earned scores of 3 or higher).

Recommendations

Based on analysis of the survey results and improved performance on AP exams, the evaluators recommended that the AP Passage program be continued. They recommended that the training for teachers be implemented a second year (the same teachers would presumably be involved) with emphasis on (a) responding to the teacher comments regarding more attention to scaffolding in the content areas of the specific courses the attending teachers taught so that all students could learn at an advanced level, and on (b) teachers' awareness of ways in which comments relating to the summer experience in class could be unwelcome to students or might even diminish self-confidence.

Because a significant portion of the invited group of students could not attend a two-week day-time program, the evaluators recommended that two options be offered. One option would be the existing schedule, and a second would be late afternoon/evening and weekend sessions spread out over the course of the summer or on a schedule that could meet the exigencies of student work schedules.

Finally, because the mindset data indicated the students entered the program with relatively high scores on the assessment of growth mindset, the measure of that variable was dropped; however, teachers were asked to be mindful of continuing to reinforce that construct.

Results of Implementation of the Recommendations

Data collected during the second year of the program indicated that the teachers perceived the training focus on the content of the particular AP courses to be beneficial; they judged the training on scaffolding to be useful; and the students did not report the concerns about being singled out by teachers in the classroom. Attendance at the AP Passage summer program increased by 10% with the addition of the second option for attendance. Scores on the AP exam were increased to a rate of 50% of the AP Passage students earning scores of 3 or higher.

CASE 3

Case 3 was created using ideas provided in Johnsen et al. (2020) and VanTassel-Baska et al. (2020) on results of practice across school districts. The school board of the Ballmore Unified School District has directed the administrators in the district to conduct an evaluation of the implementation of gifted programming. Among the goals of the program for gifted and talented students in the elementary and middle schools are:

- Professional learning activities will result in teacher mastery of understanding of the characteristics of gifted learners and appropriate strategies to modify curricular offerings and instructional strategies to match learning to those characteristics.
- Teachers will utilize appropriate differentiation strategies and other best practices in their classroom.
- Students identified as gifted and talented will demonstrate exceptional academic growth as a result of participation in the services they are offered.

Students are heterogeneously grouped with regular classroom teachers directed to deliver instruction to the gifted students in their classrooms. The school has a policy that at least two hours per week in each classroom be designated as Differentiated Instruction time.

Data

The evaluation process included collection of survey data from teachers on their perceptions of how effective the professional learning activities (three one-day summer workshops with half-day in-service activities once per month during the academic year) were in preparing them to differentiate in their classrooms. Focus group interview data were collected from teachers on which of the characteristics of gifted learners were exemplified by the identified students in their classrooms, examples of how they had modified the general education curriculum for gifted students in their classrooms, and descriptions of how they used the designated DI time each week to address differing students' learning profiles. The responses were compared to the agendas and the resources distributed during the staff development.

The building principals in all elementary and middle schools were interviewed to ascertain whether (a) they had observed differentiation in the classroom and, if so, what the examples were of differentiation, and (b) how they evaluated and provided feedback on differentiation to the teachers in their buildings.

Classrooms were also observed. Two instruments were used in the observations that occurred during both general instructional activities and during activities offered during DI time. The first, the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) provided data across three domains (Emotional Support, Classroom Organization, and Instructional Support). Classrooms were also observed using the Classroom Observation Scale-Revised (COS-R; VanTassel-Baska et al., 2003) The COS-R provides data on teacher behaviors across the domains of: Accommodations for Individual Differences, Creative Thinking Strategies, Critical Thinking Strategies, Curriculum Planning and Delivery, Problem Solving, and Research Strategies.

Students and teachers in the middle school completed the Class Activities Questionnaire (Steele, 1982) to gauge instructional emphasis across the dimensions of memory, translation, interpretation, application, analysis, synthesis, and evaluation.

Results

The data from the surveys and focus group interviews documented that the teachers had learned to identify the characteristics of gifted learners. But even though they expressed confidence that the professional development activities had provided sufficient guidance in modifying curriculum and instruction to provide appropriate challenge for gifted students in their classrooms, qualitative analysis of their examples of differentiation led to the conclusion that their level of understanding was very basic and lacked the sophistication expected for adequate responses to gifted leaners needs. The teachers also reported not knowing where to find appropriate resources to differentiate for gifted learners. These data were supported by analysis of both the COS-R and the CAQ. While the teachers and students both reported considerable emphasis on memory, translation, interpretation and application in the classroom activities, and teachers reported emphasizing analysis, synthesis, and evaluation, the students reported low levels of emphasis on the development of these skills. Further, only the areas of Curriculum Planning and Development and Accommodations for Individual Differences received ratings averaging in the "effective" range. No differences were observed between times designated as DI time and other instructional time. Field notes indicated that DI time was largely devoted to providing remedial instruction on a one-to-one basis to struggling learners, with gifted learners assigned to computer activities such as Accelerated Reader (a dubious program for advanced readers despite the title of the program). The evaluation data provided by the Classroom Assessment Scoring System affirmed teacher competence across all three dimensions of that scale.

Principals reported documenting that DI time was, in fact, scheduled in each classroom, but as a group the principals were not proficient in describing what would constitute appropriate differentiated instruction as it had been delineated in the professional learning for teachers. Furthermore, they did not report any systematic evaluation of teachers' efforts to differentiate the curricular or instructional activities for gifted learners.

Recommendations

The analysis of the data on differentiation in the classrooms suggested a need to reconsider the program delivery model and the professional development. The evaluators recommended cluster grouping gifted students with efforts made to place gifted students in the classrooms of teachers who had exhibited the greatest competency in differentiation. The professional development program should then be modified to be differentiated in itself to provide all teachers the opportunity to continue to hone those skills but providing those with emerging or higher baseline skills more advanced training. The evaluators also proposed a peer mentorship program where teachers with highly developed skills work with other teachers on developing high-level competency in differentiation. They also recommended that published units in the disciplines be provided to teachers as potential resources so that teachers would both have models of appropriately differentiated units and be able to use those units as appropriate (e.g., the problem-based learning units in science and language arts units developed at William and Mary, the units based on the CLEAR curriculum developed at the University of Virginia, and the mathematics units developed at the University of Connecticut).

Two recommendations were made relative to administrators. First, the school district leaders should develop or adapt existing scales for observation of differentiation in the classroom that would be used by principals to collect data during DI time and during general instruction on differentiation. The data suggested that principals needed professional development focusing on the principles of differentiation to appropriately implement such an instrument.

Results of Implementation of the Recommendations

During the subsequent academic year, the cluster grouping model was used to group students in both the elementary and middle schools in the district. The differentiated professional learning program was also implemented and was supplemented with a teacher-constructed reference list of resources. However, the mentoring program could not be arranged because of a lack of funding for the gifted program. Those teachers with cluster groups of gifted students each implemented at least one of the recommended units, improved their ratings on the COS-R and the CAQ, and were rated by principals as using DI time for differentiation of curriculum and instruction with gifted learners at least 50% of the time observed. However, observations by principals were not frequent enough to provide satisfactory data. Principals noted the responsibility they faced of ensuring instruction that would result in satisfactory performance on state proficiency tests as explicated by school district policy and evaluation protocols as conflicting with the requirements of using the observation scale focused on differentiation. Even though the COS-R is aligned with the Common Core State Standards and could be viewed as not in conflict, the school district policies on teacher evaluation were viewed as the process that had to receive priority in their allocation of time. Two possible remedies were proposed. One was to assign the gifted coordinator responsibility for observations using the COS-R; the second was to work with school administrators in creating a new evaluation instrument for the district that would incorporate key elements of the COS-R.

Cross-Case Elements

The three cases described above illustrate the processes and outcomes of the process of program evaluation. In each case the evaluation process focuses on the evaluation of particular components of gifted programming (e.g., identification of gifted students, professional development, curricular implementation, student outcomes), but each case also illustrates the interface between components of a gifted program or services (e.g., the relationship between the success of professional development efforts and teachers' success in implementing differentiated curriculum and instruction in their classrooms). Administrators and evaluators in these cases would have been depriving themselves of relevant and useful information for program revision (and ultimately program success) if they had limited themselves to evaluation of only one facet of the program without consideration of other potentially related and significantly influential factors.

The importance of complementary data collection strategies is also illustrated in these cases. For example, data from both the Advanced Placement and classroom grades are triangulated to support the findings of effectiveness of the Advanced Placement passage program. Had the results not been congruent (for example, students doing well in class but not on the AP exams), then other recommendations on feedback to students through grades would have been explored. Further, quantitative and qualitative data not only complement one another in the conclusions from findings; they also supplement each other and provide corresponding insights. In Case 3, quantitative data from classroom observations is complemented by qualitative data from interviews in establishing issues in the use of DI time and in the reasons why those issues existed (and suggest ways to remediate the problem with modifications in the ways students are assigned to classrooms).

Finally, the cases illustrate the tenet that both program development and program evaluation are iterative processes. That is, the best of intentions is not usually sufficient to produce expected results in the first year or two of program implementation. Further, continuing evaluation is necessary to ascertain whether or not recommended adjustments to programs are effective mechanisms to move a program closer to its goals. Descriptions of results of the execution of recommendations provide examples of both positive impacts of recommendations and failures of others.

Lessons Learned

The evaluation of program services in the cases described also offer lessons that are supported by the literature on impactful evaluation of gifted programs. These cases, in conjunction with the findings of Callahan et al. (1995); Tomlinson & Callahan (1994); and Van-Tassel-Baska et al. (2020). The literature, combined with the cases as well as and state and local level evaluations conducted by the author, provide corroboration and fundamental guidance in planning and executing evaluations with the greatest potential for positive impact.

Commit to Evaluation Process

The importance of stakeholder buy-in to the evaluation process cannot be over-estimated. First, those who are in high level decision-making positions are crucial because any resources necessary to implement change must come from the policy making, planning, and budgeting process. Each of the cases above resulted in recommendations that either require monetary resources (e.g., additional staff development, scheduling an additional AP Passage opportunity in the evenings or on weekends), necessitated a re-alignment of priorities (e.g., adding administrative responsibility to ensure that teachers are evaluated on the degree of fidelity to the proposed changes in instruction), or called for policy change (e.g., a new identification process adopted system-wide). The stakeholder commitment to the evaluation should be clear early in the evaluation process, that is, at the time the evaluation questions are determined. While sometimes the results of an evaluation can be surprising, a discussion of possible outcomes and the implications is a sound foundation for planning once the results and recommendations are presented.

A corollary to the importance of commitment is to ensure that one identifies all stakeholders who are invested in the outcomes of the various aspects of the gifted programming efforts or have an interest in or need for evaluation results and involve them in the evaluation process. The inclusion of school board members, administrators, teachers, parents of gifted students, students, etc. is critical to ensure that the issues identified are deemed important (and all-important issues are identified), that the data collected is considered credible by those who make decisions based on the data, and that the results and recommendations are transmitted to those stakeholders in ways that are useful to them in their decision-making. Callahan et al. (1995) noted that effective implementation of evaluation results is more likely when evaluation procedures are a part of planning from the earliest stages of program development and evaluation plans include a specific blueprint for the use of evaluation findings.

Be Cognizant of the Importance of Timing and Disseminate Reports to All Appropriate Audiences in a Timely Fashion

The appropriate collection and interpretation of evaluation data is highly time dependent. For example, data collected on student outcomes as a result of gifted programming in the first year of implementation of a program is likely premature as the identification processes, curriculum and instructional products and practices, teacher training, etc. are likely to evolve in the beginning years. On occasions when student outcomes are measured before teachers have had the opportunity to fully grasp the depth and complexity of a new curriculum and deliver instruction effectively, premature judgments may be rendered that a program is ineffective. However, process data such as teacher perceptions of professional learning options, classroom observation data on the degree to which a curriculum is being implemented with fidelity and the degree to which teachers are making progress toward implementing the instruction relative to the goals and objectives of the program, demographic data on the outcomes of the first iteration of implementation of an identification process, etc. can be very useful formative data to guide decision-making relative to which program components are "working" as intended and which need modification. For example, the collection of data on student attendance at the AP Passage program during its first iteration alerted the evaluators and program staff that there were likely factors inhibiting attendance that should and could be addressed.

A second important issue relates to the timing of the delivery of evaluation reports (even if they are interim) so that budgetary and scheduling options can be given appropriate consideration. In the AP Passage case study, the addition of an alternative to the summer program will require both additional funds and additional scheduling of personnel. If the recommendation comes after budgets have been finalized, the likelihood that the recommendation can be implemented diminishes considerably. Such was the situation in Case 3—funds were not available for the mentoring component of the recommendations, perceived to be important to success.

Choose or Construct Valid and Reliable Assessment Tools

Research on the impact of program evaluation in gifted education documents two important principles relative to the selection of instruments to measure either process or products of the program:

1. Develop or select assessment tools that address the complex issues of measurement that characterize outcomes of gifted programs.
2. Use a variety of data gathering methods designed to reflect the unique structure and goals of programs for gifted learners (i.e., out-of-level testing, portfolio assessment, product rating with demonstrated inter-rater reliability).

As evaluators seek to choose instruments to measure the processes and outcomes of gifted programming the push and pull between choosing instruments that are widely accepted as measures of curricular outcomes in general education such as standardized achievement tests and choosing instruments that may not be as widely known. No matter the source or type of instrument, reliability and validity are key to the production of credible results and recommendations. See pp. 270–79 (for elaboration on this point.

Use Multiple Data Sources (e.g., Teachers, Parents, Students, Administrators, School Board Members) and Multiple Assessment Tools

The validity and credibility of results and recommendations are enhanced when data can be triangulated. That is, when data from multiple data sources converge on or suggest a common interpretation of the status of a process or outcome, the result is more defensible and gives greater leverage to the recommendation. Triangulation can be based on multiple instruments used to measure the same process or outcome or can be multiple informants on the same process or product. For example, in Case 3 the results of the teacher survey and on the COS-R support the recommendations offered and are illustrative of multiple sources of data supporting the same conclusion.

However, one also must be cognizant of the burden placed on informants. When teachers are asked to fill out surveys, rate student products, and attend focus group sessions, they may become weary of involvement and not present high-quality data. Using different subgroups of each stakeholder group for different assessments can be a useful strategy to ward off assessment fatigue.

Making Decisions Based on Student Outcome Information in Isolation from Process Information May Lead to Bad Decisions

In an age of accountability, it is tempting to focus on outcomes prematurely. Note that outcome data were not collected in Case 3 even though outcomes were specified in the goal statements. The evaluators and school personnel determined during the planning stage that teachers would need time to fully integrate differentiation into their classroom practice so to expect the achievement of the specified student outcomes in the first year of the evaluation process was premature. To measure outcome data before fidelity to the curriculum and instructional model is established can lead to the conclusion that a program curriculum or model is inappropriate when it may be that the model is excellent, but the delivery is flawed.

Evaluations Are Not Controlled Research Studies

In each of the cases described in this chapter, the data that are presented as outcomes are not derived from an experimental paradigm that represents a controlled study with randomly assigned treatment and control groups. As such, the evaluators and stakeholders need to carefully consider the criteria for performance (of teachers in the differentiated classroom, of student performance on tests and performance tasks, etc.) that will be considered acceptable (or extraordinary) for performance. The determination of those criteria should be accomplished prior to the collection and analysis of data, but in the implementation of the evaluation process circumstances and context may suggest that modification for a given assessment period is needed.

Ensure Recommendations Are Structured in a Way that Encourages Follow-through

In creating evaluation recommendations, the consideration of the context for change and resources available for change are crucial. For example, while it is obvious that not all teachers without the skill or will to implement differentiation can be removed from the classroom and such a recommendation could not be implemented, cluster grouping in Case 3 was a viable option for ensuring that gifted students would be placed in classrooms where differentiation to address their learning and performance capabilities would be addressed.

Final Words

Often the consideration of evaluation is "put on the back burner" for later attention, but then it is relegated to such a low priority that it is not implemented. The descriptions provided in the case studies highlight the importance of evaluation in improving the programs and services offered to gifted students. In each of the cases we see important goals of gifted programs addressed, but then because data on the resources, the processes, and the outcomes of existing program implementation efforts guide decision-makers in improving one or more facets of their endeavors, services to gifted students are positively impacted. Diversity is increased without dilution of program outcomes, professional development is modified for greater effect on teachers' classroom practice, program services are modified to bring differentiation of curriculum and instruction into focus, and students outcomes are improved through those professional development and program modifications.

References

Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. R. (1995). *Instruments and evaluation designs used in gifted programs* (Research Monograph 95132). Storrs: University of Connecticut, The National Research Center on the Gifted and Talented.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.
Fowler, M. (1990) The diet cola test. *Science Scope*, *13*(4), 32–34.
Johnsen, S. K., Fearon-Drake, D., & Wisely, L. W. (2020) A formative evaluation of differentiation practices in elementary cluster classrooms, *Roeper Review*, *42*(3), 206–218. https//doi.org/10.1080/02783193.2020.1765921
Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-language learners using nonverbal tests. *Gifted Child Quarterly*, *52*(4), 275–296. https://doi.org/10.1177/0016986208321808
Mun R. U., Hemmler V., Langley, S. D., Ware, S., Gubbins, E. J., Callahan, C. M., McCoach, D. B., & Siegle, D. (2020). Identifying and serving English learners in gifted education: Looking back and moving forward. *Journal for the Education of the Gifted*, *43*(4), 297–335. doi:10.1177/0162353220955230
Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, *60*(3), 159–171. https://doi.org/10.1177/0016986216643165
Peters, S. J., Carter, J., & Plucker, J. A. (2020). Rethinking how we identify "gifted" students. *Phi Delta Kappan*, *102*(4), 8–13. doi:10.1177/0031721720978055
Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS*). Brookes
Renzulli, J. S., & Reis, S. M. (n.d.). Student Product Assessment Form (SPAF). https://gifted.uconn.edu/wp-content/uploads/sites/961/2015/02/spaf.pdf).
Rizza, M., & Morrison, W. F. (2007). Identifying twice-exceptional students: A toolkit for success. *TEACHING Exceptional Children, Plus*, *3(3)*. https://files.eric.ed.gov fulltext/EJ967126.pdf
Steele, J. (1982). *Class Activities Questionnaire*. Prufrock.
Tomlinson, C. A., & Callahan, C. M. (1994). Planning effective evaluations for programs for the gifted. *Roeper Review*, *17*(1), 46–51. https://doi.org/10.1080/02783199409553617
VanTassel-Baska, J., Avery, L., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). The William and Mary Classroom Observation Scales— Revised. Center for Gifted Education, The College of William and Mary, School of Education, Williamsburg, Virginia, United States.
VanTassel-Baska, J., Hubbard, G. F., & Robbins, J. I. (2020). Differentiated of instruction for gifted learners: Collated evaluative studies of teacher classroom practices. *Roeper Review*, *42*(3). 153–164. https://doi.org/10.1080/02783193.2020.1765919

Selected Resources

Evaluation of Student Products

1. *Student Product Assessment Form (SPAF)*

Renzulli, J. S., & Reis, S. M. (1991). The assessment of creative products in programs for gifted and talented students. *Gifted Child Quarterly*, *35*, 128–134. doi:10.1177/00169862910350Cla0304 (available at https://gifted.uconn.edu/wp-content/uploads/sites/961/2015/02/spaf.pdf).

2. *Creativity Demonstrated in Writing Rubric*

Kettler, T., & Bower, J. (2017). Creative capacity in gifted students: Comparing teacher ratings and student products. *Gifted Child Quarterly*, *61*(4), 290–299. doi:10.1177/0016986217722617

3. *Rubric for Secondary Level Research Projects*

Schack, G. (1994) Authentic assessment procedures for secondary students' original research. *Journal of Secondary Gifted Education*, *6*(1), 38–43.

4. *Sample Rubric Anchoring Assessment Criteria on the Definition of Creativity* and *Using the Components of Critical Thinking to Develop Assessment Criteria*

Shively, K., Stith, K. M., & Rubenstein, L. D. (2018). Measuring What Matters: Assessing Creativity, Critical Thinking, and the Design Process. *Gifted Child Today*, *41*(3), 149–158. https://doi.org/10.1177/1076217518768361

Observation of Classrooms

1. *Classroom Assessment Scoring System (CLASS)*
Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS*). Brookes. This instrument requires training in implementation. Go to: https://education.virginia.edu/classroom-assessment-scoring-system

2. *The William and Mary Classroom Observation Scale-Revised (COS-R*)
VanTassel-Baska, J., Avery, A., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). https://education.wm.edu/centers/cfge/_documents/research/athena/cosrform.pdf

Measures of Student Outcomes

1. *Diet Cola Test*
Fowler, M. (1990) The diet cola test. *Science Scope*, *13*(4), 32–34.

2. *STEM Thinking Skills Test (STEM)*
California Academic Press; information available at STEM Thinking Skills Test – Insight Assessment

3. *Iowa Assessments*
Iowa Assessments – Iowa Testing Programs – College of Education – The University of Iowa (uiowa.edu); itp.edu.uiow.edu

Classroom Climate

1. *Classroom Activities Questionnaire*
Steele, J. (1982). *Class Activities Questionnaire*. Prufrock.

Critical Thinking

1. *Test of Critical Thinking*
Bracken, B. A., Bai, W., Fithian, E., Lamprecht, S., Little, C., & Quek, C. (2004). *Test of Critical Thinking*. Williamsburg, VA: The Center for Gifted Education. Retrieved from http://cfge.wm.edu/publications.htm

Teacher Beliefs

1. *Alignment of Teacher Beliefs with NAGC (2010) Pre-K-Grade 12 Programming Standards*
Johnsen, S. K., & Kaul, C. R. (2019). Assessing teacher beliefs regarding research-based practices to improve services for gt students, *Gifted Child Today*, *42*(4), 229–239. doi:10.1177/1076217519862332

Mindset

1. *Implicit Theories of Intelligence Scale*
Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

General Source

1. *RAND Education Assessment Finder: Measuring social, emotional, and academic competencies.*
www.rand.org/pubs/tools/TL308.html

Conclusion to Part III on Evaluation of Programs

Joyce VanTassel-Baska

The chapters in this section have provided important understandings about how program evaluation works, the tools by which it can be enacted and the processes through which it can be successful. They offer a picture of what dynamic programs for the gifted might look like if program evaluation was used routinely as a tool for annual assessment of progress, both of students and the program itself. Yet recent studies (VanTassel-Baska & Hubbard, 2019; VanTassel-Baska et al., 2020) suggest that gifted programs lack fidelity of implementation in respect to inclusion of underrepresented groups, use of differentiated materials and strategies, and effective grouping practices.

Chapter 15 systematically reviews the tools commonly used in gifted program evaluations by experts in the field. The chapter provides a list of instruments that possess good technical adequacy and that have been used successfully for the purpose stated. These instruments generally have a manual to describe how they are used and where they report technical adequacy data. For quantitative instruments, inclusion of norming, content validity and reliability data, and evidence of bias testing is necessary. For qualitative measures, content validity, internal consistency, and interrater reliability should be the basic tools for judging the instrumentation's quality. Throughout the chapter, there is an adherence to the standards for evaluation, stressing the need for knowledgeable people to conduct it, the need for annual evaluations, and the need for the use of multiple tools to answer different evaluation questions. The authors have been extremely thorough in their search for quality instruments that may be (a) tests that show student growth pre-post, (b) product inventories and rating scales that assess student capability across a range of skill sets both cognitive and affective, and (c) observational tools that highlight classroom implementation variables. Each type of instrument they review is usually employed with others in a program evaluation to bolster credibility of findings through the use of multiple sources.

Chapter 16 presents to the reader a bird's-eye view of three evaluations conducted in recent years that offer case studies of individual district evaluation issues and cross-district comparisons. It provides three interesting examples of how school districts actually employ the evaluation tools discussed in the prior chapter, first to uncover the current status of the program, and then to work on an improvement agenda to make it better. The chapter reviews the core aspects of the evaluation process as

DOI: 10.4324/9781003285991-25

it is conducted, beginning with the questions asked, followed by the data examined, the recommendations rendered, and the implementation actually put in place. By examining three different sites, the reader is able to see how evaluation has value in different sizes and types of districts as well as both elementary and secondary settings. A set of lessons learned concludes the chapter to focus on the basic issues of greatest importance in program evaluation—the issues of commitment to the evaluation process itself and the need for its annual frequency, the use of technically adequate instruments, the employment of multiple sources, and the importance of timely dissemination and follow-through.

These two chapters stand as the keys to carrying out program evaluation and then knowing what to do next in program utilization. They provide a picture of what exemplary practice might look like in this area of program development. As such, they are important signposts for the future of gifted education. For it is in program improvement and remediation of ineffective practices as well as expansion of opportunities in all areas of curriculum to a broader group of learners that the field must aspire.

Yet the models that Adams and Callahan provide have not been widely used. Evaluation as a part of the program development process has never been well-established nor widely implemented in school districts. Most districts attempt to collect some data on stakeholder perceptions annually, and several have evaluations by outside consultants conducted every 5–7 years, yet others never have had their programs formally evaluated. The lack of such efforts hinders the ongoing improvement of programs which is critical to program credibility and ultimate growth and development. Part of the problem rests with the desire to believe the best about gifted programs, that they are the "holy grail" of districts' improving education for all, that such programs might bestow a level of education that could be accessed by all if given the opportunity.

The reality of gifted programs, however, is quite different. Most of them are limited in scope and comprehensiveness by content area and grade level. Most of them are under-resourced both in respect to personnel and fiscal capacity. These conditions have been reported in the literature for the past 30 years. While school districts have typically committed to annual school improvement plans, they have failed to assess the status of gifted programs as a part of that planning. Often, they have failed to implement recommended changes to gifted programs when they have been made. Even cost-effective changes have been ignored such as reducing the number of needless state-aligned benchmark assessments for gifted students and district-wide pacing guides that complicate the use of acceleration and other differentiation practices with gifted students.

Emergent Issues from Gifted Program Evaluations

There is some evidence that suggests gifted coordinators do find program evaluations helpful and informative when they are conducted and that teachers of gifted students in particular value the feedback on practice that such evaluations provide (VanTassel-Baska, 2019; VanTassel-Baska & Brown, 2022). In an earlier survey across six districts, results indicated that gifted personnel used evaluation results for three major purposes: (a) future program planning; (b) interaction with general education specialists, staff, and parents; and (c) assessment of student learning (VanTassel-Baska, 2004). Yet the nature of program changes made as a result of evaluations often are more limited than necessary for substantial program improvement. Consequently, some central issues remain for the field to grapple with as we begin a new phase of programming post COVID. These emergent issues include:

1. **The inclusion of more under-represented populations in gifted programs was perceived to be a major problem with identification systems.**

Many of the districts that have conducted evaluations were already using processes to find underrepresented populations that have been recommended by many national reports, studies, and standards for exemplary programming (see Johnsen & Kaul, 2019; Callahan & Herzberg, 2018; Plucker & Peters, 2017). While the best practice processes appear to be in operation (e.g., use of multiple measures, use of flexible cutoffs, use of student products), they do not appear to be making much difference in the results of identified students from these populations. This finding underlines the importance of doing follow-up studies to ascertain the instruments and processes that are working in situ for school districts to establish equity in gifted program membership.

2. **Primary stakeholders (e.g., students and parents) held more positive views of differentiated curriculum, even when employed unevenly, than the general education curriculum used in classrooms.**

Across the districts evaluated, both teachers and students were positively disposed toward the curriculum being implemented in the gifted program. Students found the curriculum to be "stimulating and challenging" in comparison to the general education curriculum. Teachers felt it was "more rigorous and student-centered." When parents were aware of the curriculum, their view was positive as well, often noting that it provided opportunities for their child to think and problem-solve in ways not typical from other curricular fare. Unfortunately, in many cases, their knowledge of the curriculum was not sufficiently high to rate or comment on perceptions of its effectiveness. The perception of the differences between the general education curriculum and the gifted curriculum was often not borne out by classroom observation data, however.

3. **Limited differentiation is employed in classrooms that contain gifted learners.**

Many districts use a combination of curriculum materials for their gifted programs. Where differentiated materials were used, they were employed primarily in language arts and math, commensurate with the areas in which differentiated programming was employed. In math, that curriculum was generally accelerated by one year so that gifted students were leaving middle school, having completed geometry and ready to take advanced math courses through AP Calculus in high school. In language arts, the general curriculum tended to dominate without alterations except in the use of advanced reading level materials and discussion groups. Few districts used differentiated materials in social studies or science at any level of the curriculum although STEM-based programs contributed to differentiation in science through the use of instructional inquiry processes. AP and IB programs and materials dominated high school curriculum for the gifted, offering advanced learning opportunities not found in the general curriculum.

Fidelity of implementation was observed when and where differentiated materials were employed. When teachers implemented their own units of study, however, the use of differentiation was uneven, depending on the teacher and his or her capacity to design and implement curriculum effectively.

4. **Administrator views of differentiated curriculum for gifted students were generally limited and focused on the needs of all learners for differentiation and the inclusion of more underrepresented groups in gifted programs.**

It was rather common in the evaluated districts for principals and their assistants not to know or understand what teachers were doing with differentiated curriculum. It was almost entirely perceived to be the responsibility of the teachers and the gifted coordinator to inform them of the nature of changes to the general curriculum and how those changes addressed state and national standards. Beyond that, there appeared to be a disinterest in what was being provided for their high performing students, and a tendency to shy away from offering advanced classes in the core subject areas. Most of the superintendents interviewed felt strongly about all teachers' having the capacity to use inquiry-based strategies with all learners in their classrooms. They also favored using alternative tools to include underrepresented groups in gifted programs. Some districts carefully tracked their progress in this area. Because of their concern for these two issues, superintendents tended to favor cluster grouping or in-class grouping, even when their use of more targeted instructional grouping showed evidence of growth in higher level thinking and advanced content learning for their own gifted students.

5. **The chosen type of grouping within a district influenced the use of differentiated strategies for gifted and other students.**

Although the research on grouping gifted learners tends to be supportive of most grouping approaches, with the exception of cooperative learning groups or no grouping (Rogers, 2007), districts have tended to limit the types of grouping employed in implementing gifted programs to the detriment of the use of differentiated practices. In one classroom observation study, there was a significant relationship between the grouping model employed by the district to deliver curriculum to gifted students and the use of differentiation (VanTassel-Baska et al., 2020). The more there was special class grouping at middle school and selected pullouts at elementary level, the more differentiation was observed. Cluster grouping, where it was employed, usually failed to promote within-class differentiation to any extent.

6. **The use of differentiated curriculum was viewed as potentially compromising gifted student performance on the state assessment tests in reading and math.**

Many districts employed pacing guides that were aligned to the district curriculum in very specific ways. Consequently, differentiated curriculum, designed for gifted learners, was treated as an add-on rather than as an integral part of the daily classroom routine. Alternative materials were difficult to employ as they were not aligned to the standards in the same way as the pacing guides had interpreted the district materials. Even though gifted students have demonstrated high level performance on the state content tests, they still were subjected to lower level curriculum and repetitive assessments.

These findings suggest that multipronged efforts must be enacted in order to improve the identification and related curriculum services for advanced learners in classrooms.

Multiple stakeholders, including administrators, lack understanding of instruments and the processes needed to make differentiated curriculum work. Although greater awareness and training may help, these approaches alone will not improve implementation. Only thoughtful teachers and coordinators can make sure that progress is made.

Issues related to inclusion of underrepresented groups must not dictate the structure of program delivery. Rather, the problem should alert gifted educators to the need for adjusting the bar for who is accepted into gifted programs and providing greater differentiation in those programs for students with more diverse needs. If the program is school-based, it is necessary to design more criteria for inclusion, including the extensive use of content- and performance-based data for placement.

Issues related to what constitutes differentiation for the gifted suggest that more intensive training is necessary for both teachers and administrators on gifted student curriculum standards, effective instructional differentiation and materials, and alignment to content standards. The trend of interpreting differentiation as individualization, tailoring every lesson to each student's individual need, must be supported by more efficient approaches to implementing differentiation.

While the use of inquiry models like problem-based learning clearly motivates many students, it is not a panacea for enhancing advanced learning in core content for students who do not have prerequisite skills. Thus, project-based approaches for group implementation must be carefully designed to account for differences in learning rates and levels.

Finally, issues related to the limited use of strategies and materials found to be effective with gifted populations suggests the need for more monitoring of classrooms and student progress by administrators who must be trained in how to evaluate the nature of instructional intervention and its efficacy for top learners.

Conclusion

Gifted students should be served in program models that work, delivered by trained teachers through pacing that reflects their readiness to learn and through advanced content that challenges them. Curriculum that matches the strengths, needs, and interests should be used with all learners. However, not all learners benefit equally from above-level content, a faster pace, and problem-based learning. Decisions about instructional approach should always be made, based on assessments of individual student's readiness to learn, using underlying skills as a building block to providing advanced work and to pursuing their interests independently. The greatest overall benefit of gifted programs to general education might be to provide inquiry-based instructional approaches for all students while offering the necessary advanced opportunities to gifted learners through a variety of instructional grouping models. Evaluation allows us to verify and validate those benefits for all.

References

Callahan, C., & Herzberg, H. (Eds.). (2018). *Fundamentals of gifted education (2nd ed.).* Routledge.

Johnsen, S. K., & Kaul, C. R. (2019). Assessing teacher beliefs regarding research-based practices to improve services for GT students. *Gifted Child Today*, *42*(4), 229–239. https://doi.org/10.1177/1076217519862332

Plucker, J. A., & Peters, S. J. (2017). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Educational Press.

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly*, *51*(4), 382–396. https://doi.org/10.1177/0016986207306324

VanTassel-Baska, J. (2004). Metafindings on the utilization of evaluations. In J. VanTassel-Baska & A. Feng (Eds.), *Designing and utilizing evaluations for gifted program improvement* (pp. 155–174) Prufrock.

VanTassel-Baska, J. (2019). The role of evaluation in gifted program development: Coordinators' portraits of progress. *Gifted Child Today*, *42*(4), 240–248. https://doi.org/10.1177/1076217519862323

VanTassel-Baska, J., & Hubbard, G. F. (2019). A review of the national gifted standards implementation in eight districts: An uneven picture of practice. *Gifted Child Today*, *42*(4), 215–228. https://doi.org/10.1177/1076217519862336

VanTassel-Baska, J. & Brown, E. (2022) An analysis of stakeholder perceptions of gifted programs: A Report card on gifted program performance. *Gifted Child Today, 43*(3).

VanTassel-Baska, J., Hubbard, G. F., & Robbins, J. I. (2020). Differentiation of instruction for gifted learners: Collated evaluative studies of teacher classroom practices. *Roeper Review*, *42*(3), 153–164. https://doi.org/10.1080/02783193.2020.1765919

CHAPTER

17

Issues and Insights on the Process of Assessment

Susan K. Johnsen and Joyce VanTassel-Baska

Major Principles from the Chapters

In our review of the chapters in the book, certain concepts about assessment emerged, creating a viable set of principles that applied to all three applications of the term assessment that we have used in the book. These principles were also supported by the authors' current views about the topic and the research literature.

Principle 1. Assessments Should Be Selected Based on Their Technical Qualities and Their Alignment to Advanced Student Characteristics and Specific Gifted Services and Programs

When the general education program is not addressing the strengths and needs of students, they need to be referred for specialized services and programs. As all of the authors noted, tests should be selected based on their technical qualities and alignment to the gifted program's specific goals and outcomes. For example, if the school has developed accelerated programs for students who are above level, then assessments need to be able to show students' advanced progress and performances. Moreover, these assessments need to have norms (e.g., How do students who are above level perform on the assessment?), be reliable (e.g., Are observations of students' above-level performance similar across classrooms? Is the assessment standardized?), and be valid for the purpose (e.g., Does the assessment have above-level items and/or show important characteristics of students who are performing above level?). Both qualitative and quantitative assessments need to meet technical adequacy standards (see Chapter 4, Little et al., p. 72). For example, rating scales can provide both types of information. As Ryser mentions, "Standardized and well-designed rating scales can result in more valuable and consistent referrals from teachers and/or parents" (Chapter 5, p. 89). Fairness or equity is also determined by the technical aspects of the test (see Chapter 5, p. 90). Technical manuals should provide information related to accommodations that might remove construct-irrelevant barriers (e.g., English directions for English language learners), inappropriate uses (e.g., identifying general reasoning vs. specific academic skills),

DOI: 10.4324/9781003285991-26

subgroup differences (e.g., students from lower socio-economic vs. those from higher socio-economic backgrounds), and even items that might discriminate against specific students (e.g., rural vs. urban; see Chapter 3, Lakin et al., p. 50; Chapter 5, Ryser, p. 90; Chapter 6, VanTassel-Baska & Johnsen, p. 119).

This principle applies equally to all the tools for assessment mentioned by authors in Part II on assessing learning. Both technical adequacy of the instruments and their alignment to standards for gifted curriculum and their subject area dominate the discussions of authors in this section. Even as state testing measures are decried as inadequate for assessing advanced levels of curriculum, they are still recommended for use in conjunction with performance and product-based tools to provide a broader portrait of gifted student learning (see Brown, Sulak et al., Kettler & Lamb). Moreover, each of the authors in this section also lament the inadequate evidence of effectiveness of many of the existing performance-based tools (see Rinn & Walker, Moon et al., Sulak et al.), suggesting that their use may be limited to individual classrooms rather than being generally applied.

In respect to evaluation, the third section of the book, the theme of the Adams and Caughey chapter clearly focuses on the need for technical adequacy in the selection of tools used to judge the efficacy of programs and services while also promoting the selection of instruments appropriate for use with advanced learners. Encased within most evaluations are the data used to document student growth, the second aspect of assessment we have presented. Moreover, most evaluations examine the development processes employed in a program, including identification. Thus, evaluation as assessment encompasses all of the areas of assessment of interest in this book.

Principle 2. Since Assessments Provide Different Types of Information, Multiple Assessments Should Be Used

Assessments provide different types of information. Lakin et al. explain the Cattell-Horn-Carroll model of intelligence that includes a continuum of assessments ranging from measuring fluid reasoning (i.e., general ability) to crystalized intelligence (i.e., school-acquired knowledge). Quantitative assessments might measure general ability using intelligence tests whereas achievement tests might be used to measure school-acquired knowledge. Products, performance-based and other qualitative assessments are particularly helpful in revealing a complex range of behaviors, which show the student's readiness for relevant gifted service and program opportunities (see Chapter 4, Little et al., p. 73). In addition, classroom-based evidence and specific performance assessment tasks engage the student, provide representative samples of students' work over time that show learner progress, increase equitable access by recognizing talent potential, and are often better predictors of student potential and future performance. Therefore, all the chapter authors agree that multiple criteria improve the selection process and need to be used for identifying underrepresented groups. No single test solves equity challenges (see Chapter 2, Lee & Peters, p. 39).

Learning assessments, as those used for identification, require multiple tools in making high stakes decisions about the learning that gifted students have mastered. While programs like AP and IB do not require such evidence for decisions for credit or placement at the college level since coursework is calibrated to first year college work, K-12 education often does require more evidence to support advanced placement and credit. The use of standardized measures coupled with performance-based measures make the most convincing tools for such a purpose. The use of state assessments and content-based measures used in combination provide convincing evidence of student readiness for advanced level or more complex work within a domain (see Chapter 12, Rinn & Walker; Chapter 8, Sulak et al.; and Chapter 13, Boswell et al.). Teacher and parent observations also are useful in adding to the evidence base of data supporting students' readiness and their interests particularly in domains that may be outside of the school setting.

Evaluations also benefit from a combination of multiple qualitative and quantitative assessments. Making judgments about programs and services for gifted students clearly require the use of multiple measures to make decisions about their effectiveness. Since evaluations probe the program systems such as identification and learning progress, the use of data sources such as focus groups and program operation monitoring allow evaluators to understand how these processes function. Since evaluations probe the outcomes of advanced learning opportunities, pre-post measures of student growth need to be analyzed. Finally, since evaluations also probe an understanding of a variety of programming and multiple services as exemplars of best practice in the field of gifted education, the use of different evaluation tools need to be employed.

Principle 3. When Assessments Are Administered to Every Student (i.e., Universal Screening), Students from Underrepresented Populations Are More Likely to be Identified and Served

Lee and Peters in Chapter 2 describe how universal screening works to improve the sensitivity and the equity of identification systems by improving nomination validity, reliability, and subjective screening or referral criteria. They emphasize that quantitative tests are not biased in and of themselves, but rather how they are used. For example, test manuals can show how they minimize bias by reporting scores for different subgroups and examining individual items to see if they are biased against particular groups (i.e., differential item functioning). Qualitative tools used during the referral process should be based on observing behaviors, not vague descriptions, and should be combined with many observations in different contexts that complement quantitative measures. Single-phase, universal systems that use quality instruments will allow a larger number of students to be served; however, if using a two-phase system, which reduce costs, the phase-one cut scores need to be lowered (60th to 90th percentile) for cutoff purposes.

The interpretation of this principle for learning assessments rests in the need for access to advanced programming for eligible students and access to advanced assessments as requested. Written information on the criteria for selection into the program and expectations for performance should be open and transparent. In the case of AP and IB, programs and assessments align and are elective for all learners (Brown). Yet the learners who opt for these programs tend to be those encouraged to take them, those who meet prerequisites, and those who score at advanced levels on relevant assessments of aptitude. Many of these students, however, are not identified gifted students in some school districts, raising the question as to why they were not selected for programs at earlier stages of development. It may be due to prior years' education that has prepared them for advanced coursework in specific domains or for performance on specific aptitude versus more general ability measures. Of those students from underrepresented groups who participate in both the programs and concomitant assessments, success has been reported to be high (College Board, 2019).

Other learning assessments often have a looser connection to learning outcomes. This problem creates a need for multiple assessments to demonstrate proficiency in a given area of the curriculum. The use of common advanced content assessments at elementary and middle school levels, such as the Fowler test of scientific investigation, can benefit all students by providing consistent data for teachers to examine skills central to standards across levels and facilitate learning in areas where students' errors have been discerned (see Chapter 12, Rinn & Walker). Writing assessments may also be used on a universal basis for judging progress in this critical area of the curriculum (see Chapter 7, Kettler & Lamb). While access to advanced programming renders a greater inclusion of underrepresented groups, it does not guarantee their readiness to learn at advanced levels, making the use of assessments for differentiation of the curriculum critical.

In evaluation, the determination of the effectiveness of systems of program management is partially based on the extent to which equity is a cornerstone of identification and program opportunities. Thus, assessing the extent to which equity is a standard of program functioning is a part of the evaluation process. Also central to an effective evaluation system is the implementation of the goals and outcomes of the program. Lack of program fidelity is a common problem in school programs and services that impede effectiveness (see VanTassel-Baska & Brown, 2022).

Principle 4. Assessments Should Be Selected, Administered, and Interpreted by Qualified Individuals

To interpret assessments, educators need to understand different types of tests (e.g., norm- vs. criterion-referenced), comparable scores (e.g., raw, percentiles, index), measurement error, and how to combine information from qualitative and quantitative assessments (see Chapter 3, Lakin et al.; Chapter 5, Ryser). Are we comparing the scores to a national population or local sample? Are we using a common scale? Are we considering that the true score lies within a range of scores because of the error in the test? Are we combining qualitative information and quantitative information to determine each student's interests, strengths, and needs? Each member of the selection committee needs to be well-versed in the answers to these questions.

Assessment tools for learning also require expertise in selection, scoring, and interpretation of results. Many assessments that are standardized already have technical manuals and/or publisher resources that describe these processes or individuals with expertise to carry out these tasks although the expertise in interpretation of data for gifted programs may be overlooked in favor a norm-based perspective. School psychologists and counselors have expertise in assessment tools and their appropriateness for particular populations like the gifted; gifted educators have the expertise in the characteristics and needs of the gifted students. These two groups are recommended for inclusion on assessment and placement committees that determine placement of gifted students who need specific services and programs.

Evaluation tools demand the expertise of educators of gifted students in the selection of appropriate tools and the interpretation of their results. Because triangulation of results is a central tenet in interpreting evaluation data, experts in both evaluation and gifted education need to be involved in the interpretation of findings and the implications drawn for recommendations. Moreover, evaluation findings should be studied for the extent of implementation by school districts and their continued viability.

A Model of Assessment in Gifted Education

In examining the different purposes for assessments in gifted education and the common principles that are shared, we explored the relationships across each type of assessment related to their goals, common processes, and instruments.

Identification is an ongoing process of gathering information by using tests, instruments, and techniques that match the characteristics of gifted students and programming opportunities (see Chapter 1, Johnsen & VanTassel-Baska, pp. 21–22). The process is intended to encourage all students to express their gifts and talents so that educators can find those students who need services and programs. Given the importance of the identification processes for meeting the interests, strengths, and needs of gifted students, the chapter authors paid particular attention to pre-identification activities, types and qualities of assessments, implementation and interpretation, and equity issues.

Learning progress is an ongoing dynamic system of growth experienced by gifted learners as they traverse the opportunities provided them through both formal and informal programs and services. These assessments in and of themselves provide new growth experiences that provide appropriate indications of the level of functioning of individual learners on their journey of talent development. They also provide school stakeholders (i.e., teachers and administrators) the evidence for advanced placement and continued opportunities for more rigorous, complex, and creative learning. In these chapters, the authors introduced alternative measures that may aid in that process.

Evaluation is also an ongoing process with the goal of continual program improvement. Gifted programs and services frame outcomes that represent advanced opportunities for learning in one or more areas at all stages of development. Students selected for these opportunities are chosen through a valid, reliable, and equitable process that creates an optimal match between identification and programming. Evaluation then uses multiple data sources to determine the extent to which these processes were employed and to the extent to which they were effective in respect to student learning and benefits.

Figure 17.1 portrays these interrelationships of the three aspects of assessment in real time for program development. Based on student characteristics, the first level incorporates assessments related to pre-identification development and the formal

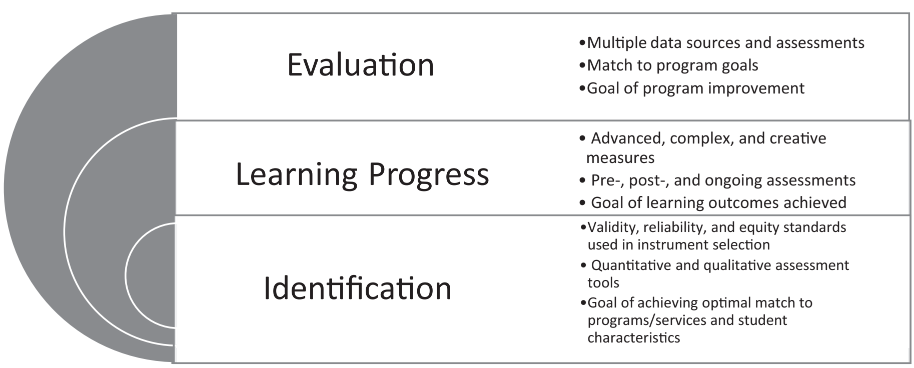

FIGURE 17.1 The interplay of identification, learning progress, and evaluation

identification cycle. Pre-identification includes any learning experiences that provide opportunities for students' talents to emerge (e.g., talent development pools, flexible grouping with high ability students, advanced learning activities). These pre-identification activities are particularly helpful for students from underrepresented groups who may not have access to educational opportunities. From these activities, from differentiated learning activities in the general education programs, and from referrals will emerge students who need to be screened further either in a one-phase or two-phase system (i.e., universal screening where all students receive assessments or where there is initial screening and then further assessments). During these phases, assessments are aligned to both the student's characteristics and the services and programs. Following the collection of information during the identification phases, a committee reviews the information and places students in services that are aligned to their interests, strengths, and needs. Learning progress assessments continue to be used to examine how successful the students are in the programs and their retention. These data are also helpful in determining which of the assessments are particularly useful in predicting performance. Program placement also sets up the dynamic by which the program and services may be monitored for fidelity of implementation, a keystone for ensuring that a goal-based approach to evaluation may be effected.

Practical Implications Emerging From the Principles and the Assessment Model

When examining the principles and the intersections across different types of assessments, educators need to consider a variety of approaches for implementing identification, learning progress, and evaluation assessments.

Practical Implication 1. State and Local Policies Affect Implementation of any Assessment Process

In their survey of directors of gifted programs, VanTassel-Baska and Johnsen in Chapter 6 noted that three states had made changes (or were in the process of making changes) in state policy to accommodate efforts to increase the number of underrepresented students identified for gifted programs (p. 109). As an example, Texas has included this accountability standard in its plan: "2.25. The population of the gifted/talented services program is closely reflective of the population of the total district and/or campus" (Texas Education Agency, 2019, p. 13). Obviously, for districts in this state, directors will want to align their district policies and identification procedures so that equity is achieved. Other states are now also changing policies to account for issues of equity in their identification policies.

Practical Implication 2. Effective Assessment Systems Consider the Student Population, Phases of Identification and Development of Programs, and How Multiple Assessments Can Be Used to Enhance All Aspects of Assessment, Particularly for Students from Underrepresented Groups

Given that both identification and learning assessments include both quantitative and qualitative measures, the process of learning about the strengths and needs of gifted students is ongoing (see Chapter 1, VanTassel-Baska & Johnsen, p. 23). Districts need to decide which quantitative assessments will be administered annually during formal identification and learning outcome processes. Will the administration include a one- or two-phase system prior to selection (see Chapter 2, Lee & Peters)? While district benchmark tests may provide some information each semester, teachers continually collect information using differentiated learning experiences about student progress and performances, which in turn, can be used in referring students for additional assessments. Portfolios can also be constructed and used as part of the case for decision-makers in identifying or providing evidence of advanced learning (see Chapter 4, Little et al., pp. 76–77). Moreover, developmental talent pools can emerge from universal screening efforts for students showing potential so that more students are engaged in advanced curriculum or challenging response lesson opportunities, which may lead to a broadening of the identification process and an adaptation in programs (see Chapter 4, Little et al., pp. 75–76). All of these components, of course, require collaboration across teachers, administrators, parents, and other community stakeholders.

Practical Implication 3. Educators Need to Consider the Purposes and Types of Assessments to Administer

Knowing that the use of multiple tests and multiple sources is superior to a single test score and the judgment of just one teacher, educators need to carefully identify assessments (see Chapter 5, Ryser, p. 88). The selection needs to be based on the assessment's purpose as well the characteristics of the students and the gifted education program and services. Cognitive ability tests may be used to reflect potential for advanced learning (see Chapter 3, Lakin et al., p. 44). If the program focuses on acceleration in specific domains, above-level or cross-grade level achievement tests will be needed. If rating scales are used, sources of information (e.g., teachers, parents, counselors, peers) may vary to collect information in different contexts (see Chapter 5, Ryser, p. 88). Moreover, English learners may require nonverbal assessments; twice exceptional students, multi subtest assessments to reflect uneven performances; and students from poverty, performance-based measures to demonstrate complex thinking and growth over time (see Chapter 6, VanTassel-Baska & Johnsen, pp. 101–105). Effective combinational approaches work together to yield students who require gifted education services and who are from underrepresented groups. They also work in providing a multidimensional portrait of student learning as districts assess learning progress.

Practical Implication 4. To Interpret Assessments for Any Purpose (i.e., Identification, Learning Progress, Evaluation of Programs), Educators Need to Understand the Test's Purpose, the Norming Population and Subgroup Scores, the Test's Scores, the Standard Error of Measurement, and Variations in Student Performance across Assessments

The chapter authors agree that educators need to understand that different types of tests have different purposes. Some students' performances on tests may relate to one another (e.g., achievement tests to cognitive ability tests) whereas others may not (e.g., higher portfolio scores than standardized achievement test scores). These differences may relate to the validity and reliability of the assessments, perhaps the norming group (i.e., national norms or local campus norms), or intraindividual characteristics (e.g., students with disabilities, students from poverty, English learners). When comparing scores, educators need to always consider the standard error of measurement, which is based on the reliability and dispersion of scores (i.e., standard deviation) (see Chapter 3, Lakin et al., pp. 48–49; Chapter 5, Ryser, pp. 92–93). The true score generally lies within a range of performance so that strict cut off scores should not be applied. Educators also need to consider validity studies with subgroups. How do subgroups perform on the test? Given the variation in student performance across assessments, profiles or matrices should not be collapsed into a single score (see Chapter 3, Lakin et al., p. 60). Educators should thoughtfully examine evidence and consider indicators that show developed skills, readiness, and potential in both the identification and learning outcome documentation process.

Practical Implication 5. Professional learning is important for teachers in understanding characteristics of gifted students, gathering performances from the classroom, administering and interpreting assessments, and employing differentiated classroom practices. Professional learning is also important for parents in understanding their gifted students, the learning outcomes and delivery of the gifted education programs and services in their district, the identification process, and the importance of advocacy.

The process of identification, whether in one or more stages, needs to be conducted by educators who have received professional learning in gifted education (see Chapter 4, Little et al., p. 74; Conclusion to Part I, VanTassel-Baska & Johnsen, p. 119). The authors identified misconceptions influencing teacher referrals and their effectiveness in the identification process. For example, teachers may believe that giftedness is synonymous with high achievers and not examine variations in performance across domains (see Chapter 3, Lakin et al., p. 57). They may also have low teacher expectations for some groups, particularly students from poverty and students of color, and not refer these students based on implicit biases (see Chapter 5, Ryser, p. 95). For these reasons, teachers need to participate in professional learning to dispel myths, to understand the characteristics of gifted and talented students, and to understand how tests are influenced by student background and prior educational experiences. When prepared, teachers are more likely to refer a more diverse pool of students (see Chapter 4, Little et al., p. 79). If not, then teachers overlook talent in students who do not fit traditional definitions. Educators also need to understand the technical qualities of tests, learn how to administer tests, and how to gather performances from the classroom. Finally, if they serve on selection committees, they need to learn how to examine multiple assessments in making decisions. Such needs are also prevalent in the use of learning progress assessments (see Chapter 8, Sulak et al.; Chapter 12, Rinn & Walker; Chapter 7, Kettler & Lamb). Moreover, administrators also need to gain expertise in the assessment processes for gifted learners. While they may not be personally involved in the relevant selection, scoring, or interpretation committees, they do need to understand their purpose and desired outcomes to the extent that they can lend support to program administrators at the building level.

Practical Implication 6. School District Resources Need to Be Sufficient to Include the Costs for All Three Types of Assessments, Educator Time, Professional Learning, and Monitoring the Fidelity of Each Process

Most of the chapter authors examined costs in terms of assessments, time needed for implementing the assessments, and monitoring the fidelity of the identification process. Group-administered tests were considered more feasible and cost-effective for universal screening than individually administered tests (see Chapter 3, Lakin et al., p. 54). Lee and Peters (Chapter 2) suggested that two-phase systems, when implemented correctly, can generate similar sensitivity to a one-phase, universal system at far less cost. The first-phase screener should be easier, faster, and cheaper than phase-two assessments that determine service eligibility. In addition, costs of assessments can be minimized by examining their alignment with one another (Lee & Peters, p. 33). If two assessments appear to be measuring the same construct, one can be eliminated.

Although curriculum-based assessment tasks are likely to be less expensive than published standardized assessments, they may add costs with regard to personnel resources (see Chapter 4, Little et al., p. 74). Educators who are charged with administering and implementing performance-based tasks will need the time, training, and resources to accomplish these efforts with fidelity. Corroborating Little et al.'s comments, VanTassel-Baska and Johnsen reported that several coordinators mentioned that teachers and they themselves had to put in many extra hours in order to ensure that the identification system was implemented with fidelity (p. 109). Coordinators recognized that this was not a one-year issue, but rather one that would require multiple years of using multiple approaches. Professional learning and the allocation of time and resources appear to be critical.

Identified Challenges

Educators should also expect to encounter challenges, even as they adopt best practices in all three forms of assessment. These challenges have been commented on in several chapters in this book but deserve to be highlighted here in the final chapter.

Challenge 1. No Assessment System Assures Equity

All authors agree that equity is a challenge. Structural inequities exist that include personal, economic, social, and institutional (Chapter 3, Lakin et al., p. 62) factors. Multiple approaches must be employed to make progress on identifying students from underrepresented groups (see Chapter 6, VanTassel-Baska & Johnsen, p. 110). These include front-loading talent development activities, universal screening, using locally developed norms, assuring assessment tools are in the child's preferred language for communication or nonverbal formats, and building relationships with students (see Conclusion to Part I, VanTassel-Baska & Johnsen, pp. 118–20; Programming Standard 2.3.1, National Association for Gifted Children, 2019).

Challenge 2. Intra- and Inter-individual Student Differences in the School's Population May Exclude Some Gifted Students

Students may be missed because they have strengths in some areas and weaknesses in another, depending on the subgroup. For example, twice-exceptional students' gifts may mask their disabilities and vice versa; English learners may not perform well on assessments with high verbal demands; students from poverty may not have had access to educational opportunities; and students of color may experience implicit bias and racism (see Chapter 3, Lakin et al., p. 53; Chapter 6, VanTassel-Baska & Johnsen, p. 100). Moreover, all of these factors interrelate, increasing the complexity of equitable identification. In some cases, districts focus on finding only particular underrepresent groups (e.g., Black and Latinx) and ignore others (e.g., 2E and those from lower income backgrounds), which leads to additional problems of proportionality (Chapter 6, VanTassel-Baska & Johnsen, pp. 105–106). Differences of "between group needs" as well as "within group (individual) needs" require attention. It is important to tailor identification and services to the strengths, needs, and interests of underserved and under-resourced groups of students.

Challenge 3. Giftedness is influenced by access and educational advantages

There is a misconception that cognitive ability tests measure innate capacity (Chapter 3, Lakin et al., p. 56). Many differences in performance can be attributed to educational advantages, which result from access to quality educational opportunities and pre-K programs. As children matriculate in school, the achievement gaps increase for lower and higher income children (Chapter 6, VanTassel-Baska & Johnsen, p. 106). Moreover, educators need to use pre- and ongoing assessments to differentiate the curriculum for the increasing diversity in gifted education programs.

Challenge 4. Qualitative Assessments Have Less Technical Data Supporting Their Use

While performance tasks may increase the number of students served from underrepresented populations, they have less technical data supporting their use than standardized measures. Therefore, assessments need to be balanced, including authentic representations of students' classroom work with more standardized, reliable measures (see Chapter 4, Little et al., p. 81). This balance is also true in assessing learning progress of students who are receiving gifted programs and services and in determining the overall effectiveness of the program.

Challenge 5. Assessment of Gifted Students Requires Material and Human Resources

Standardized, quantitative assessments have strong technical data but can be expensive, particularly when administered to all students (see Chapter 2, Lee & Peters, p. 59). Curriculum-based assessment tasks, which are less reliable, may be less expensive but require more training and administrative time (see Chapter 4, Little et al., p. 74). Balancing the two types of assessments needs to be an administrative consideration at both identification and learning progress levels. Moreover, ongoing program evaluation may require the expertise of external evaluators who can provide different perspectives related to overall gifted program and system improvement.

Challenge 6. Educators Have Misconceptions about Gifted Students

Educator misconceptions are influenced by the dominant culture, their experiences, knowledge of gifted characteristics, and the child's status (e.g., limited opportunities to learn as a result of poverty; physical or learning disabilities; fluency with English). These misconceptions then influence the validity and reliability of qualitative instruments (see Chapter 4, Little et al., p. 72). For example, teacher ratings of classroom behaviors may not accurately reflect the student's behaviors due to rater bias (see Chapter 5, Ryser, p. 94) or students' lack of knowledge about behavioral expectations. Teachers must therefore have extensive training in using checklists and other nomination forms for identifying gifted students from underrepresented populations. They also will need professional learning in implementing differentiated and advanced learning experiences for gifted students who have different levels of readiness, interests, strengths, cultural backgrounds, and needs.

Challenge 7. Research Over Time Is Required to Determine the Effectiveness of All Three Types of Assessment Systems (Identification, Learning Progress, and Evaluation) in Order to Make Needed Modifications

The identification process needs to be continuously examined to ensure that it is effective, efficient, and implemented with fidelity. Which assessments are effective in identifying students who are successful in the program? Which assessments identify students from underrepresented groups? How do the identification tools align with the programs and services? Which learning assessments are both efficient and effective—requiring less educator time but highlighting advanced learning in programs? What do talent trajectories look like for advanced learners in each subject area? After data are collected, educators need to disseminate the results and use the data to improve future identification and learning cycles and optimize instructional services (see Conclusion to Part II, Learning Progress, VanTassel-Baska & Johnsen, p. 256). Details of the outcomes from changing assessment practices must become part of a written record to disseminate to school boards and advisory groups in the district to ensure fidelity of implementation, continuity, and improvement of efforts.

Future Research Questions

While a robust literature exists for selecting and using assessment tools for identification, less is known about the use of alternative measures for learning assessments. Questions still remain that may be addressed through new research. Some of these questions are noted as follows by the type of assessment to be studied.

Identification

1. What is the effect of pre-identification programs on improving proportionality (i.e., what is the predictive validity of using these programs)?
2. What existing performance-based assessments have enough ceiling to use in the identification of gifted students?
3. What is the sensitivity, the cost efficiency, and equity of each of the assessments used in the identification process? How can these three criteria best be balanced?
4. What is the effectiveness of one-phase and two-phase systems in identifying students from underrepresented groups?

5. What types of assessments (i.e., quantitative and qualitative) best predict successful performance in specific programs? How might they be combined?
6. What is the relationship between different cut-off scores and successful performance in programs?
7. What aspects of the identification process are related to the long-term retention of students in gifted education programs?
8. How does early identification relate to closing achievement gaps?
9. Which assessments are most predictive of career and general life outcomes?

Learning Progress

1. What performance-based measures are most predictive of attaining gifted learner outcomes?
2. How can qualitative measures' validity and reliability be improved to measure learning progress?
3. How might a standardized design format improve the construction and effectiveness of qualitative assessments and rubrics?
4. What model of professional learning is most effective for conveying to teachers the principles and practices for learning progress assessments?
5. What assessment model for advanced learning is most cost-effective?
6. What mechanisms might schools employ to make gifted assessment of learning easier to accomplish and more meaningful for ongoing curriculum development and implementation for all learners?

Evaluation

1. What evaluation data sources yield the most important findings for program improvement?
2. How important is triangulation of perspectives of stakeholders for program improvement?
3. What processes are most effective for synthesizing findings across data sources?
4. What barriers restrict program recommendations from being implemented? How might those barriers be overcome?
5. What dissemination plan would be most effective in communicating evaluation findings?

Conclusion

This book has been about the role of assessment in gifted programming. Its three sections have demonstrated the importance of having (a) an identification system that is valid, reliable, and equitable, (b) a learning progress system that employs different types of tools that uncover advanced growth patterns in gifted learners, and (c) an evaluation system that employs multiple data sources to answer specific evaluation questions about the effectiveness of gifted programs. The book has featured authors with specific expertise in the types of assessments being explicated from university researchers to program and instrument developers to school practitioners. It has laid out the research evidence available for using different assessment tools and recommended processes and procedures for use. Finally, this chapter has illustrated how the assessment systems interact and share common principles and implications for best practice.

References

Callahan, C. M., Moon, T. R., & Oh, S. (2013). *Status of elementary gifted programs*. Charlottesville, VA: National Research Center on the Gifted and Talented University of Virginia. http://www.nagc.org/sites/default/files/key%20reports/ELEM%20school%20GT%20Survey%20Report.pdf

National Association for Gifted Children (2019*). 2019 Pre-k-grade 12 gifted programming standards*. www.nagc.org/sites/default/files/standards/Intro%202019%20Programming%20Standards%281%29.pdf

Texas Education Agency (2019). Texas state plan for the education of gifted/talented students. https://tea.texas.gov/sites/default/files/GT_State_Plan_2019_1.pdf

VanTassel-Baska, J. & Brown, E. (2022) An analysis of stakeholder perceptions of gifted programs: A Report card on gifted program performance. *Gifted Child Today, 43*(3).

Index