CHAPTER 21

METHODS TO INCREASE THE IDENTIFICATION RATE OF STUDENTS FROM TRADITIONALLY UNDERREPRESENTED POPULATIONS FOR GIFTED SERVICES

Michael S. Matthews and Scott J. Peters

Alternative models for identifying students for gifted programs have been a focus of study since at least the early 1970s, when substantial scholarly attention first began to be directed toward the lack of diversity among gifted students. In recent decades, researchers worked to develop identification procedures that recognize a broader range of talent, and that are presumed to be less influenced by students' cultural, linguistic, and economic background. These efforts can be broadly categorized as advocating for the use of different or alternative measures (i.e., using different tests) or for a different application of traditional measures (i.e., using the same tests differently). The former approach includes the use of nonverbal ability tests (e.g., the Test of Nonverbal Intelligence [TONI]) and structured observation protocols (e.g., U-STARS+), whereas the latter approach includes methods such as universal screening, local, and group-specific norms for traditional aptitude tests or systems for combining and including multiple measures.

With roots back to the pioneering efforts of psychometricians like Martin Jenkins (1936), there has been a growing awareness of the lack of equity in educational outcomes, particularly those measured by standardized assessments (i.e., differences in mean scores for different groups). In K–12 settings, observable outcomes of these score differences include differential representation rates in special education (but see Morgan et al., 2015, for an argument disputing differential representation in special education) and gifted education programs, as well as differential rates in college attendance, degree completion, and many other important outcomes. This relative lack of diversity in cultural, racial, and economic terms is evident within the population of students identified by schools to participate in gifted programs (Yoon & Gentry, 2009), as well as in the excellence gaps in student achievement documented by Plucker, Burroughs, and Song (2010).

Whether it is framed in terms of equity and excellence, social justice, economics, or even as a desire for strict numerical proportionality, identification is a crucial issue, and one over which there has been much gnashing of teeth. There seems to be a collective desire for greater proportionality of representation, and yet there is little agreement on how to go about achieving that end. Additionally, differential representation—the widespread observation that gifted classrooms do not look like a representative sample of the larger general school populations—is likely a key reason why gifted education has been subject to a great deal of criticism from other disciplines (e.g., sociology; Margolin, 1994). In this chapter, we outline contemporary efforts to create and implement "alternative" models of gifted education identification or program selection. We address where these models have been successful and where they have stumbled, and on the basis of these findings, we offer recommendations for practice for psychologists and teachers who are charged with developing and implementing identification practices in K–12 schools.

http://dx.doi.org/10.1037/0000038-021
APA Handbook of Giftedness and Talent, S. I. Pfeiffer (Editor-in-Chief)

IMPORTANCE OF THE TOPIC

Racial and sometimes ethnic differences are visually salient; add this fact together with the troubled history of race relations in the United States and it should not be surprising that racial differences in identification rates of students for gifted education have received the greatest attention. Some authorities have gone as far as to assert flatly that "underrepresentation [in gifted programs] is an issue of race and racism—period" (Dickson, 2015, para. 2). Yet, other differences, including disability status, culture, language, and socioeconomic status (SES), are also highly correlated with lower gifted program participation rates. So, while undeniably important, race clearly is not the only relevant issue. For the field to advance, it must move beyond a singular conception of diversity toward a more nuanced and inclusive view that extends its scope to the myriad additional ways in which individuals are different from and similar to one another. If the field truly believes that "outstanding talents are present in children and youth from all cultural groups, across all economic strata, and in all areas of human endeavor" (Ross, 1993, p. 3), then more needs to be done to make proportional representation in gifted and advanced programming a reality.

Despite some notable efforts (e.g., Baldwin & Wooster, 1977; Bernal & Reyna, 1974), after more than 40 years of study, there still seems to be relatively few research-based best practices to guide psychologists, teachers, and other practitioners in developing and implementing identification procedures in the schools. When there is research available it often conflicts with other research findings, leaving practitioners confused or frustrated. Continually evolving perspectives about the nature of giftedness (e.g., Dai, 2009; Kornilov, Tan, Elliott, Sternberg, & Grigorenko, 2012; Pfeiffer, 2012) also complicate efforts to ameliorate underrepresentation in gifted programming.

In this chapter, we summarize some alternative identification approaches that have been proposed to make gifted education more inclusive, and explain the limitations of these approaches so that readers may make well-informed choices about identification practices. In addition, we offer some of our own suggestions—which are, in turn, based in other scholars' work—for how we might proceed in addressing this important issue. But first, a caveat bears mentioning. There is a simple but very important rule that should guide any attempt at more inclusive gifted identification: If a school district is going to alter its identification practices from what is already in place, it must also alter the services it provides to identified students (assuming that there was already strong agreement between identification and services). Regardless of which version or combination of the following methods a district chooses to implement, any change to identification practices will require some expansion or modification of the resulting services offered to students. A district cannot simply change how students are admitted to a program, and then still expect that program to meet the now-diversified needs of the identified students. For example, if a district begins giving a new test to all students in an attempt to locate more English language learners, a gifted population will be identified that is different from those who were identified last year (which was, of course, the reason for adopting the alternative assessment). This means additional or altered services will be necessary to meet the needs of some members of the newly identified group. It is never appropriate to simply add a new identification pathway while ignoring what effect this will have on the success of the resulting programs, services, and interventions.

Traits, States, and Educational Gaps

There seems to be broad agreement in the field that underrepresentation is a problem and that one of the goals of contemporary gifted education should be to address this underrepresentation in some shape or form (Erwin & Worrell, 2012; Matthews & Kirsch, 2011). Despite the salience of race, growing attention also is being devoted to cultural and linguistic inequalities in identification, as well as to economic disparities' role in educational achievement (see, e.g., the Jack Kent Cooke Foundation [http://www.jkcf.org]). Though a few educational gaps do appear to have narrowed over the past decade (Plucker, Hardesty, & Burroughs, 2013), very little attention has been paid to narrowing gaps at the advanced end of the achievement spectrum. Growing media

attention to this issue and favorable legislative changes offer some reasons for hope that this situation will be improving; reporting provisions in the Every Student Succeeds Act may provide some of the accountability tools needed to more widely address gaps in the performance of high-ability learners in U.S. schools.

Regardless of what can be said about the success of closing minimum proficiency gaps over the last 20 years, Plucker and colleagues showed that excellence gaps have grown during the post-No Child Left Behind (2002) era. If equity is a concern, then attention must also be paid to equity at the highest levels of performance, in addition to the substantial attention already being focused on bringing up achievement at the lower end of the spectrum. Some research supports the idea that addressing economic and linguistic disparities can also narrow observed racial/ethnic disparities, likely due to the overlap in membership among these different categories (cf. McBee, Shaunessy, & Matthews, 2012). Later in this chapter we offer some suggestions specifically related to the identification of ability among students whose first language is other than English, as well as some cautions that should be kept in mind.

One of the reasons for persistent disparities in participation rates may be traced to the ongoing lack of consensus regarding how giftedness should be defined. The trait view of giftedness (see Dai, 2009) asserts that giftedness is an inherent quality of the individual, and that once possessed, giftedness persists over the lifetime of the individual regardless of current educational needs, current level of challenge, later accomplishment, or anything else. In this view giftedness is eternal and permanent and its status as a trait within a person is the reason specialized services are needed.

The trait view of giftedness is evidenced particularly regarding affective characteristics. For example, the National Association for Gifted Children (n.d.) says that although every learner is different, "there are many traits that gifted individuals have in common," and lists various traits characteristic of gifted individuals, including sensitivity or empathy to the feelings of others, keen sense of humor, and intellectual curiosity (Clark, 2008). Although such lists provide a starting point for discussion, traits alone are an insufficient basis for defining a multifaceted construct like giftedness (Pfeiffer, 2012). It is also worth knowing we were unable to find much in the way of empirical sources to support the claim that these characteristics were found only in gifted learners and not in bright or even average learners, or that these traits are in any way exclusive to those with high academic or intellectual ability.

When giftedness is viewed as a trait of people, then the goal of gifted identification is to seek out those who have that trait and sort them from those who do not. There are many problems with this approach, but the most salient for this discussion is that this model has been practiced for decades, while at the same time drastic rates of underrepresentation have been observed for African American, Latino, American Indian, and low-SES students. An outsider looking at this state of affairs might justifiably wonder whether children from these populations are inherently inferior in some way, or at least whether they must be somehow inherently less gifted.

Another undesirable outcome of the trait view is that gifted programming looks as if it is designed only for wealthy, Caucasian, or Asian students because those are the only children who "have" the trait of giftedness. As we said, there are many more problems with the trait conception of giftedness, but it is clear that following the trait view for many years has yielded drastic underrepresentation and led to pervasive low levels of support for gifted education (meaning few if any increases in public desire for, or state support of, expanded services).

An alternative to the permanent trait view of giftedness is to view giftedness as a temporary state. In the state view, giftedness is understood as a temporary state of unmet need—the reason students are labeled or identified as gifted in this perspective is because they currently have an educational need that is not being met. The key advantages of this point of view are twofold. First, it is intensely local in time and space. Rather than permanently branding an individual as being gifted across all contexts—a judgment based on imperfect evidence collected at a single point in time—the state view of giftedness simply observes that the individual is capable of more advanced work than that with which he or she is currently being provided.

This mismatch between readiness and curriculum/instruction is the definition of need. Second, the issue of disproportional representation becomes much less of an obstacle when local comparisons, such as those within a community or a classroom, are the basis for this judgment. Instead of searching for those few people who were born with the trait of giftedness, we are instead looking for those students who are being under-challenged by their current educational placement and instruction at a specific point in time. Lack of challenge is much more measurable than a nebulous and permanent trait known as "giftedness." With current identification practices that use national norms, some schools may have no gifted students identified (a position we find preposterous), whereas others (usually in more affluent settings) may find that nearly all of their students can be identified as gifted. We believe the larger purpose of gifted education is to nurture and develop the skills and talents of students within a local educational setting. Using such a philosophy, there are always students—in every school and in every population—who could be doing more.

Two difficulties inherent to the state view are its susceptibility to biases and its relatively messy nature in implementation. In terms of biases, if giftedness must be determined from the outside—as by a teacher or parent who notices an unmet need in some way—then any biases these individuals may hold can become reasons for incorrect identification decisions (see Peterson, 1999; Siegle, Moore, Mann, & Wilson, 2010); though, it is worth noting this is also a problem with the trait view and its related identification processes. Imprecision in criteria feeds into this as well; without clearly defined criteria, which in turn should be used because of their relationship to positive educational outcomes in advanced settings, we cannot say who would benefit from such programming simply by using a label such as gifted.

One further objection we have heard raised against the state view is that it eliminates the need for the gifted label (see related discussion in Matthews, Ritchotte, & Jolly, 2014) or even for the concept of giftedness altogether—and those who make this objection fear that without a label, gifted programs may no longer receive the limited attention in schools that they currently attract. However, we suggest that on the contrary, under the state view every school would need to offer gifted programs for those learners who are capable of more advanced work than others in their setting. There would no longer be buildings with zero percent identified gifted, because there are unmet needs in every educational context. This would be an improvement over the current status of gifted programs, in which some schools (usually those serving low-SES or traditionally underrepresented student populations) have no gifted programs at all—because under the prevailing trait view, they have no gifted learners. We believe that implementing the state view would result in a much larger population of identified gifted students being served than are currently served by following the trait view, even under the most liberal of definitions.

The Role of Intelligence Testing

Ever since Lewis Terman's (1925) pioneering work in gifted education, the identification of giftedness has been tied closely to intelligence testing, and specifically to the IQ score. Though intelligence tests are very effective in terms of low false positive rates (e.g., few people are accidentally given an IQ score of 130 if their true IQ is 110), their drawback lies in having a far higher rate of false negatives in terms of determining whether that child needs additional challenge. In other words, children are unlikely to be incorrectly identified as gifted based on an IQ score metric, but many children are likely to be incorrectly identified as not gifted using this approach. Remember, we are referring to gifted as having a need for an educational service not currently being provided.

How likely is incorrect identification to happen? Recent simulation work by McBee, Peters, and Waterman (2014) provides an estimate. In a single-assessment system, using a test with a very high reliability coefficient of .95 and a 90th percentile cutoff for identification, the probability of a correct identification (i.e., sensitivity, or correctly identifying a gifted student as gifted) is just .843. This means that under the most ideal contexts, the least messy identification systems will still miss about 15% of the students that should be identified. Additionally,

if the cutoff score is set higher, which it almost always is, the error rate increases. In other words, "even under the best condition . . . more than 15% of gifted students would be missed by the identification system" (p. 75)—and these are unrealistically ideal conditions. Under different combination rules, and with measures having lower reliability or lower intercorrelations (e.g., group tests or teacher rating scales), it is possible to devise a system in which fewer than half of students in the gifted range are identified as being eligible to receive gifted services. This is following typical identification procedures where the goal is to "find the gifted", who are people with the trait of giftedness. Clearly, these disheartening figures suggest that there may be room for other approaches.

This has led many in the field of gifted education to embrace "multiple measures" as a near mantra in terms of a best practice. The implication is that if only multiple measures were used, the problems of underrepresentation or high false negative rates would be mitigated. Under the same conditions as described previously (two tests with .95 reliability, 90th percentile cut scores, and intercorrelations of .80), a multiple measures system in which students are admitted if they meet either of the two criteria (an "or" combination rule), only 13% of gifted students would be missed. However, 15% of the identified population would have been identified incorrectly (i.e., false positives). What this demonstrates is the key point that any identification system must seek to balance false positives with false negatives. As criteria are raised and more tests are required (an "and" combination rule), false positives decrease, whereas false negatives increase. Alternatively, as multiple potential pathways are allowed ("or" combination rule) and thresholds are lowered, false negatives go down, whereas false positives go up. For this reason, allowing multiple pathways to identification might appear appealing, as it will result in fewer students being missed; but the tradeoff for increasing the size and relative diversity of the population is that the size of the population which was incorrectly identified also will increase dramatically.

Given the disagreement within the field about what giftedness is, what it really means, how it should be assessed and used in schools, and the host of other challenges, identification is complicated even in the simplest of cases. In the following sections, we outline a range of methods that have been offered in the scholarly literature as possible solutions to the problem of underrepresentation. Some have also been presented as overall better methods of identification, but we will focus our discussion on the underrepresentation aspect. In presenting these efforts we categorize them into two broad categories: those that advocate for the use of entirely different assessments and identification processes (i.e., use different tests), and those that advocate for different applications of traditional measures (i.e., use tests differently).

APPROACH ONE: USE DIFFERENT TESTS

Selecting Alternative Measures

The basis for determining which specific tests and assessments might be discarded in favor of others is not always as clear as it should be. A variety of tests have been developed to locate larger numbers of traditionally underrepresented learners. However, it is not clear that underrepresentation itself is an indication of faulty or invalid tests. Most of the alternative assessments we detail in this chapter fail to articulate what it is about existing assessments, besides the fact that certain student populations tend to receive lower scores, that indicates they are faulty and in need of replacement. Although a complete discussion of test bias is beyond the scope of this chapter, it is important to note that by itself, the observation of differential performance in the form of mean score differences by a particular group is insufficient to establish that a test is biased. Just because the boys in a science class score lower, on average, than the girls, it does not mean that the cause of that lower performance was gender bias. Bias only exists in the statistical sense of the term when two students who should have scored the same (because they had the same "true score") did not. The question, then, for gifted education is how to know when two students should have scored the same and, absent that information, what else besides bias could have caused the observed differential performance. The simple answer is that there are many

other explanations of observed score differences between racial, ethnic, or income groups that are far more plausible than test bias. Readers interested in a more complex understanding of test bias and fairness should see Camilli (2006).

There have been numerous efforts to find the one perfect test that is not influenced by extraneous factors (e.g., family income), and that, ideally, would result in proportional representation within populations they identify as gifted. A great deal of resources are (and continue to be) devoted to this search for the perfect measure. Needless to say, such a measure has not been found, though these efforts have led to some positive changes in testing in other respects. In the effort of full disclosure, we believe this desire to find or develop a one perfect test is a lost cause and is a distraction from more fruitful efforts to address the problem. There is no such thing as a culture-free test. American students are simply far too diverse, have too varied a preschool learning experience, and go to schools that vary far too widely in terms of their quality for all of these students to score the same (on average) simply because they are in the same grade.

Nonverbal Assessments

One approach within the use-different-tests category that has long been proposed in the scholarly literature was the use of nonverbal-format tests of intelligence or ability, including measures such as the TONI–4 (Brown, Sherbenou, & Johnsen, 2010), Comprehensive Test of Nonverbal Intelligence (Hammill, Pearson, & Wiederholt, 2009), the Naglieri Nonverbal Ability Test (NNAT–2; Naglieri, 2008), and the Ravens Progressive Matrices (Raven, Raven, & Court, 1938/2000).

One of these measures, the NNAT, has been heavily promoted as being able to identify proportional numbers of traditionally underrepresented students to participate in gifted programs (Naglieri & Ford, 2003), and is often described as "culturally neutral" in its assessment of ability. However, Lohman (2005a) suggested this claim was dubious at best and only could have been achieved through a reweighting of the data—something any test could do to obtain proportional representation, as this procedure is similar to a group-specific norm or preference. At least one other recent study (Giessman, Gambrell, & Stebbins, 2013) also has examined this measure's performance, and concluded that "gifted programs should not assume that using a figural screening test such as the NNAT–2, without other adjustments to selection protocol, will address minority underrepresentation" (p. 101). Overall, this study found that proportional representation was not achieved by using such measures. These findings were similar to those reported by Carman and Taylor (2010). The fact that we know of no published research that has demonstrated that such measures can eliminate or even significantly mitigate the problem of underrepresentation suggests that the NNAT and other nonverbal-format measures are not the panacea that some have claimed them to be.

The other substantial difficulty with simply changing over to nonverbal identification measures, or in using them as one of multiple pathways, also comes back to the match between identification and services. The majority of gifted programs, in our experience, places heavy demands on students' verbal and linguistic abilities in English. If students were to be identified using nonverbal and nearly language-free assessments, then students placed in a program in this manner would have little to no mastery of the instructional medium. In a perfect example of our golden rule, gifted programs would have to be changed from current practice to align more closely with the spatial abilities and figural logic that constitute the areas of ability primarily measured by nonverbal-format tests. Additionally, it is not clear what a curriculum based in spatial and figural abilities might look like, though it is possible it might have more hands-on and vocational content (cf. Gentry, Hu, Peters, & Rizza, 2008). Additionally, such a curriculum might not align well with current curriculum standards or frameworks, and these may not be sufficiently flexible to allow schools to offer such programming. We want to point out that there are schools that have developed nonverbal-specific gifted programs that align well with the skills and abilities assessed by nonverbal ability tests (cf. Peters, Matthews, McBee, & McCoach, 2014). However, in our experience, these are not the programs for which nonverbal ability

tests are most often used to identify students. Most often students are identified using nonverbal-format tests and then placed in "standard," often language-heavy, gifted programs. Even if such applications were successful in reducing underrepresentation, there is no evidence that the students identified are in need of or would be successful in the resulting programs and services, and in fact some evidence (Lohman, 2005b) suggests these approaches may actually make the problem worse by identifying a subset of traditionally underrepresented students who would not be those in their group most likely to benefit from these existing gifted programs. Clearly, more research is needed.

Native-Language Assessments

Much of the argument in favor of nonverbal ability assessment deals with the fact that performance on such tests is not influenced by English language proficiency. Another way to accomplish this same goal is simply to administer a given assessment to the student in his or her native language, as opposed to in English. For example, some academic ability or achievement tests are available in Spanish, which is the majority second language in U.S. schools. These could be given to native Spanish speakers to avoid biasing their performance results because of differing English proficiency levels. However, as was also a challenge with nonverbal ability tests, many of the programs to be provided to those successfully identified as gifted will likely be conducted in English. If students are identified using one language but then provided services in another, we could be setting students up for failure in the resulting program. Using native-language assessment also assumes that students are proficient enough in their native language to complete the test in such a way as to yield valid data. Given the wide variability in English proficiency among American students, similar variation in native language proficiency is likely to present a significant barrier to using native language assessment. Other likely bottlenecks include a lack of staff proficiency in most languages other than English that may be represented among the student body, and the lack of psychometrically validated translations of appropriate measures in most of these other languages. Nevertheless, when possible, native language assessment can help educators determine if students' performance on a test given in English was compromised by language differences by comparing their performance on the English and native language version of the same test.

Grades

Grades can indicate high achievement when they are based on evaluation of a process or product relevant to the advanced curriculum in a gifted education program (Frasier, Garcia, & Passow, 1995). Grades' advantages include their widespread availability, the lens they provide into individuals' developmental history over time, and their basis in individual performance across a variety of related tasks. For better or worse, they also include a range of skills not considered by more standardized assessments (e.g., the ability to navigate school and the classroom) that can actually be predictive of program success. However, grades may be based in part on characteristics that are not relevant to the identification of high ability (e.g., doing extra work, being cooperative, having good behavior, conforming to gender stereotype roles). Additionally, grading standards vary across teachers and schools, sometimes dramatically, making comparisons across settings problematic. Grades, likely, are an appropriate indicator of giftedness only to the extent to which the context in which they were received matches closely to the context of the gifted or advanced programs being offered; for example, students whose grades are at the top of the distribution in a math class likely would be good candidates for enrichment or acceleration in math (see also the related discussion in Pfeiffer, 2015).

Student Products and Portfolios

Student products have the potential to offer a superior means of identifying high ability, provided they represent the same sorts of performance that students identified as gifted or academically advanced will be expected to complete in the program for which they are being selected; we emphasize, again, the importance of identification criteria and program alignment. This is because prior performance is the best indicator of future performance. Tests of general ability, in contrast, can be farther removed

from future performance because they might not match as closely to these future activities (an all too common problem), and this makes them less effective as predictors.

A portfolio is simply a collection of an individual's prior work in a domain, usually chosen by the individual to exemplify their best work, which can be evaluated in some way by others. For tasks whose outcomes are not disputed and are also readily observable, portfolios are the most effective approach to evaluating the relative performance of different individuals. For example, in domains as diverse as sculpture, surgery, or kitchen remodeling, a collection of performers' best prior work will allow others to decide who they want to hire for future work in these areas.

The primary reasons portfolios are not used more widely are their cost (in time, which translates into money) and the difficulty of standardizing portfolio evaluation procedures across settings (viewed as another aspect of their higher cost). This difficulty in standardizing evaluation procedures for portfolios also results in sometimes devastatingly low or unknown reliability (Sweygers, Soetewey, Meeus, Struyf, & Pieters, 2009). With children whose work is being evaluated to determine if they might benefit from placement in a more advanced learning environment, another salient concern may be how to ensure that children have completed the work in the portfolio without undue assistance from someone else. Of course, this concern is not limited to children, as varied scandals in collegiate athletics (e.g., Garcia-Navarro, 2015), government agencies, and scholarly publishing (e.g., Barbash, 2015) have demonstrated. One way to address this concern may be to also collect samples of student work under controlled, observable conditions, and to evaluate these products for consistency with each applicant's other portfolio content. This is similar to what the U-STARs (Harradine, Coleman, & Winn, 2014) and Young Scholars (Horn, 2015) programs attempt to do. In doing this, it should be recognized that work completed on the spot may not be as high quality as work carefully selected from among a larger body of products, though it should bear a substantial similarity.

One way to make portfolio assessment less costly, in terms of additional time required, is to integrate the selection of portfolio work with existing classroom products that already are being completed anyway. In this way, every student has the opportunity to produce a portfolio, and the school population can be evaluated systematically for gifted program placement. Portfolios also can be shared with parents during student-led conferences, making them a multiuse effort that will help teachers and students to see their value beyond the gifted identification process.

Ongoing professional development related to how to evaluate portfolio products is a key need if portfolios are to be implemented successfully (Challis, 1999). Specifically, we have observed in practice that many rubrics for student products focus primarily on the presence or absence of specific features, rather than on the quality of the content or thinking that they demonstrate. For example, a rubric might award full points for using at least four different colors in a picture, or for including a topic sentence, three examples, and a closing sentence in a paper. Although such rubrics may streamline the grading process, they do not encourage bright students to put their best efforts into a project, because these learners realize that mechanics and satisfactory effort—rather than creativity, exemplary effort, or other evidence of higher-level learning—is being rewarded. Consulting lists of higher-level thinking, such as Bloom's Cognitive Taxonomy, and holding thoughtful discussions with teaching colleagues during planning time allocated by administrators for this purpose, can help in addressing this issue.

Structured Observation Protocols

What might be considered the most promising and "cutting edge" identification systems under the use-different-tests category are those which provide students with a task, lesson, or even a full unit of study, and then observe which students learn it the fastest, grab on to the most complex components, perform best, or seem to take to the style of learning. The two best examples we have seen of such systems are the Project U-STARs observation protocols (Harradine et al., 2014) and the Young Scholars program (Horn, 2015). In both cases students are presented with learning contexts that are structured, somewhat standardized, and presented to an entire

group of students. In some cases, this takes the form of a standard instructional lesson. In this fashion, such systems reduce the problem of conflating student opportunity or experience with aptitude or ability to learn, because everyone is provided with the same material at the same time. Teachers using U-STARs reported feeling like the system helped them identify more underrepresented learners, but specific data about students' subsequent success in whatever program they were being identified for was less clear and data about actual identification rates were not included. Similarly, Horn (2015) reported a significant increase in the number of identified students from underrepresented populations, and also provided additional anecdotal evidence as to those students' success in the resulting programs and services. These approaches are similar to the trial placement strategy advocated by Peters et al. (2014), but with the added benefit that they can be conducted prior to actual placement and thereby reduce the disruptions to classroom environment and uncertainties in staff allocation that result from moving students across settings partway through the school term. Though their application currently is limited to specific instructional settings, we believe this is an area worthy of further development and additional study.

APPROACH TWO: USE TESTS DIFFERENTLY

If using different tests is not the clear answer, perhaps existing tests can be used in a different way to better locate diverse learners for advanced programs and services. This approach is closer to what many advocate as current best practice in gifted identification (Lohman, 2009; Peters & Gentry, 2012).

Using Local Norms

Whereas most standardized measures are based on a nationally representative (or at least nationally collected) sample, local norms are those developed on the basis of the performance of children in a local sample—often within a district, school, or even a single grade level within a single school. Although national percentiles and comparisons are useful for some purposes (such as college admission or competition for national scholarships), they are far less useful in setting cutoff scores for admission to school-based programs, because they say nothing about the match between need and service. For example, knowing a child received an IQ score in the 99th percentile tells us nothing about whether that child is currently underchallenged in school. Local norms, in contrast, can yield information about which students are most likely to be under challenged. This information regarding which students in a school's population are most likely to need (and benefit from) advanced interventions is of the most interest to teachers and parents. As a fortuitous side effect, local norms also can be designed to yield a gifted population that is more representative of the overall school population, though with some important caveats.

We wish to emphasize an important point about local norms—one that we believe is crystal clear and is subject to little debate or disagreement. Local norms are almost always preferable to national norms for gifted identification. Unless a program or school serves a population that is drawn from the entire nation or state, there is no benefit to using national or state-level norms. Once a school decides which measures to use for identification, on the basis of the nature of the program to be provided, its identification criteria, cut scores, or percentiles should be established or set on the basis of local norms. To reiterate, aside from making comparisons of student achievement across settings, there is rarely any educational benefit to the use of national norms for making identification decisions related to local programs and services. For this reason, we believe that nearly any test or set of tests that a district decides to use for gifted identification should also use local norms.

Setting Cut-Off Scores

How, then, should cutoff scores be established? The simple answer is that cut-off scores or entrance criteria should be based on the level of the measured construct necessary for students to be successful in the program. In other words, if a student with an IQ of 130 will be successful in the program, but a student with a 129 will not, then that's where the cut-off score should be set. Unfortunately, the situation

is rarely so clear cut as this example might suggest, but the principle still holds. The main basis on which any and all identification criteria should be set are whether they predict success in the program to be provided. If the program is not closely based on the identification criteria, then these criteria are far less likely to predict student success despite the use of local norms. So, programs also should include systematic efforts to monitor student success and to connect it back to an examination of the degree of effectiveness of their identification criteria. Without this important step—and it is one that is only rarely attended to—it is difficult or impossible to justify the existence of the program.

Beyond this general rule there are two primary considerations, one theoretical and the other practical. First, any cutoff score should set a reasonable balance between the goal of being as inclusive as possible and minimizing the likelihood of admitting students whose chances of achieving success in gifted programming are relatively low. This is known as balancing false positives and false negatives. This decision depends, in turn, on three factors: the extent to which the admission decision is high-stakes (in terms of being difficult to reverse, or of leading to unfavorable programmatic consequences if the decision later has to be reversed), the consequences of failure for the individual being placed (in terms of his or her future performance and success, including affective outcomes), and the chances that overinclusion of marginally qualified learners may dilute the quality of the programming being provided, in terms of content, pace, or other characteristics that have made it suited for advanced learners in the first place.

A second important consideration is the number of spaces available in the program. A strict cut-off score can be selected on the basis of rank ordering of scores that will fill an exact number of seats; however, this primarily norm-referenced approach will create variability, sometimes wide variability, in what students selected are actually able to do. It could also end up excluding very needy students simply because of space limitations. Alternatively, a criterion-referenced approach, which evaluates applicants' performance on specific tasks, will provide a closer match to skills needed in a given program but may yield differing numbers of participants each year depending on the particular pool of skills present in those participating in the assessment. For example, selecting the top 5% of 100 students on the basis of a measure of advanced skill in math will yield a pool of five students. Now, suppose the class first is divided into five equal-sized group based on some criterion, such as height, that is mostly unrelated to mathematics performance. Now, choosing the top student on the math measure from within each of the five height groups will still give us five students, but they will be distributed more equitably across the overall group on the measure of height. This may come at the expense of not identifying a few taller students from another height group who might have had higher math scores, possibly because they are also older. This change in the selection process means we also will need to change the programming being provided to these learners so selected.

Choosing instead all students from the same group of 100 applicants whose performance demonstrates mastery of basic geometry proofs might yield five students, or twenty-five, or none, but this process likely would offer a much more coherent group of mathematics learning needs in the identified student population.

In practice, depending on the number of spaces available and the need for mastery of prior content, a hybrid of these two approaches may offer a better solution than either does alone. If the ranking approach is further narrowed to within-group rankings, it will be guaranteed, by definition, to identify representative numbers of students from within each of the groups, but likely the overall population selected will then be even more heterogeneous in terms of mastery of specific content-related skills. This also suggests the need for a hybrid of the two approaches.

Combining Multiple Measures

Multiple-criteria assessment procedures often are recommended as best practice in gifted identification (Krisel & Cowan, 1997; Matthews, 2015; McBee et al., 2014; McCallum et al., 2013; Pfeiffer, 2012). One relatively well-known drawback to multiple-criteria identification systems is that, in

many cases, they ultimately boil down to an IQ score. Renzulli and Gaesser (2015) called this the multiple criteria smokescreen, observing that "what appears to be a multiple criteria approach ends up being a smokescreen for a more traditional cut-off score approach" (p. 93). Care should be taken in constructing rules for combining scores to avoid this pitfall.

Even when they are not used as a smokescreen, multiple measures have other limitations. Depending on how they are used, considered, and combined, their outcome can range from helpful to devastating.

Thoughtful attention to the extent to which the measures to be used are intercorrelated, how strong the validity and reliability are for each measure, and especially the goals of the process and how these inform the methods of combining the scores is needed to develop procedures that minimize unintended consequences (McBee, Peters, & Miller, 2016; McBee et al., 2014). In essence, there is a continuum between inclusion on one end, and exclusivity on the other. Procedures that favor inclusion—specifically, allowing qualifying scores on any measure (the "or" combination rule), rather than requiring them on each of a set of measures—will also lead to the selection of a population whose abilities and other characteristics are more heterogeneous. Conversely, requiring cut-off scores on each of a set of measures, which McBee and colleagues labeled the "and" rule, leads to a much smaller population whose abilities are very homogeneous. Therefore, the choice of combination rules has implications for the size of the population that will be identified, and for how similar the identified students will be to one another. These in turn have implications for administration, staffing, and budgets, and for the curriculum and instruction to be provided.

Establishing Decision Criteria

Establishing appropriate decision criteria for access to gifted education programs is only necessary when there are a limited number of spaces in a program (something we believe should be far less common) or when an incorrect decision to include a student could potentially be harmful; thus, decision criteria are not necessary when all interested students can be accommodated, and in fact they may be a waste of resources in such scenarios. If space is less of a concern and allowing anyone who is interested to participate will not result in any harm to the students or the program, then identification criteria are needlessly labor intensive. We believe the field of gifted education spends far too much time and resources implementing identification systems that are needlessly exclusionary, when in many cases it would be more appropriate to simply provide these programs for all interested learners.

Additionally, decision criteria should be, but often are not in practice, tied closely to specific program content. For example, a measure of math proficiency, no matter how well developed it may be, is likely a poor way to assess whether students should be allowed access to a gifted history class. Similarly, reading proficiency would not be a good indicator of students' potential for success in a Lego-based enrichment program, though we have heard of this happening in practice (Peters et al., 2014).

OUR APPROACH: ADVANCED ACADEMICS

One of the difficulties inherent in either of the test-based approaches described in this chapter is that they still assume that testing and formal gifted identification are necessary to serve as a gatekeeper to accelerated, enriched, or otherwise advanced programs or learning opportunities. The argument behind this idea seems to be that if we just let everyone or even simply more people in, there would be dire consequences. The likely worry seems to be that advanced programming would have to be reduced in rigor, or dumbed down, to let more students complete it successfully. In fact, this has been observed to some extent in Advanced Placement coursework, as enrollments in these courses have increased dramatically over the past few decades, yet passing rates on some advanced placement exams have remained low despite increases in the number of students participating (College Board, 2009). However, it is increasingly being recognized that persistence is important, in addition to ability. Though it does make scheduling difficult, schools should strive to make it easier

for individuals to move into (and out of) different levels of coursework so that students can achieve an appropriate fit, without unnecessarily restricting the course or classroom placement opportunities for highly motivated students who may not have mastered all prerequisite skills.

It is true that advanced students benefit from being in homogeneously grouped classroom settings, for a variety of reasons (Nomi, 2009; Vogl & Preckel, 2014). These include the fact that in such settings they are less likely to be bored and can work at the faster pace of which they are capable of learning, without being held back by other learners who may need additional repetition or more background information before they can master the same content. Additionally, teachers are more effective when they can concentrate their instruction on a narrower range of student ability, which is what homogeneous grouping provides. However, neither of these very good reasons requires that a test to be used as a gatekeeper. In fact, doing so risks confusing the test itself for what it is designed to predict, which is success in a particular educational setting. Actual success in the setting is a far better predictor of future success than a less closely related test can provide. This approach can be seen in a college major (e.g., engineering) that initially attracts large numbers of interested students. Anyone admitted to college can take a freshman engineering course, but only those who are successful in this course can go on to complete an engineering degree. Additionally, noncognitive factors such as persistence or grit, self-efficacy, and achievement orientation also have a lot to do with success in these types of challenging endeavors (see, e.g., Fredricks, Alfeld, & Eccles, 2010). Although these noncognitive factors may not be captured by the point estimate of ability that a test provides, a trial period or similar approach can be effective for identifying students who possess these additional motivational and affective characteristics that will enable success in a more advanced educational setting despite these individuals' lower test scores. In simply lowering the initial barriers to participation in an advanced intervention we are also likely to find larger numbers of students from traditionally underrepresented populations.

SUMMARY AND CONCLUSIONS

Unfortunately there is no magic bullet. Ultimately the selection of alternative identification practices and criteria should be determined on the basis of a combination of factors that include the availability of spaces in a program, the prerequisite skills required for a student to be successful, the difficulty of potentially reversing a placement decision if it turns out not to be a good fit for a particular individual, and the degree to which the student is motivated to struggle, work through, and persist to achieve success in a more challenging and rigorous educational placement. By considering each of these areas, keeping in mind the key points we have discussed previously, alternative decision models can provide an effective means for matching students to appropriate learning experiences.

The simplest recommendation we can give to a concerned practitioner about how to go about locating larger numbers of underrepresented learners is to first consider why these students are not being identified with traditional measures (often because they have had fewer educational opportunities to learn, and, as such, have received lower scores on measures of the related skills—see Ferguson, 2008). Many of the previously described interventions will increase the number of underrepresented students identified to some extent. However, simply increasing numbers is not a victory on its own. There is also the goal of placing students in programs that help them learn. No matter what a district does to reach better representation rates, something must also be done to help assure newly identified students are successful. In many cases, we have observed no systematic efforts to evaluate the success of gifted programs with any students, much less with those identified through alternative or nontraditional pathways. As stakeholders engage in discussions about how to better serve low-SES and minority students, hard questions need to be asked about what the goal is, where students are starting out, what interventions would help them close that gap, and whether these approaches are in fact successful in fostering this goal.

References

Baldwin, A., & Wooster, J. (1977). *Baldwin identification matrix inservice kit for the identification of gifted and talented students*. Wooster, MA: DOK.

Barbash, F. (2015, March 27). Major publisher retracts 43 scientific papers amid wider fake peer-review scandal. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/morning-mix/wp/2015/03/27/fabricated-peer-reviews-prompt-scientific-journal-to-retract-43-papers-systematic-scheme-may-affect-other-journals

Bernal, E. M., Jr., & Reyna, J. (1974). *Analysis of giftedness in Mexican American children and design of a prototype identification instrument*. Austin, TX: Southwest Educational Development Lab.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence, fourth edition (TONI–4)*. Austin, TX: Pro-Ed.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Westport, CT: American Council on Education/Praeger.

Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri nonverbal ability test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, *54*, 75–84. http://dx.doi.org/10.1177/0016986209355976

Challis, M. (1999). AMEE medical education guide no. 11 (revised): Portfolio-based learning and assessment in medical education. *Medical Teacher*, *21*, 370–386. http://dx.doi.org/10.1080/01421599979310

Clark, B. (2008). *Growing up gifted: Developing the potential of children at home and at school* (7th ed.). Boston, MA: Pearson/Merrill Prentice Hall.

College Board. (2009). *The fifth annual AP report to the nation*. New York, NY: Author.

Dai, D. Y. (2009). Essential tensions surrounding the concept of giftedness. In L. V. Shavinina (Ed.), *The international handbook on giftedness* (pp. 39–80). http://dx.doi.org/10.1007/978-1-4020-6162-2_3

Dickson, K. (2015, December 4). The underrepresentation issue [Electronic mailing list message]. Retrieved from http://www.apa.org/ed/schools/gifted

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*, 74–87. http://dx.doi.org/10.1177/0734282911428197

Ferguson, R. F. (2008). *Towards excellence with equity: An emerging vision for closing the achievement gap*. Cambridge, MA: Harvard Education Press.

Frasier, M. M., Garcia, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students*. Storrs, CT: National Research Center on the Gifted and Talented.

Fredricks, J. A., Alfeld, C., & Eccles, J. (2010). Developing and fostering passion in academic and nonacademic domains. *Gifted Child Quarterly*, *54*, 18–30. http://dx.doi.org/10.1177/0016986209352683

Garcia-Navarro, L. (Host). (2015, January 1). Four UNC–Chapel Hill employees out in wake of cheating scandal [radio broadcast]. Retrieved from http://www.npr.org/2015/01/01/374417450/four-unc-chapel-hill-employees-out-in-wake-of-cheating-scandal

Gentry, M., Hu, S., Peters, S. J., & Rizza, M. G. (2008). Talented students in an exemplary career and technical education school: A qualitative inquiry. *Gifted Child Quarterly*, *52*, 183–198. http://dx.doi.org/10.1177/0016986208319300

Giessman, J. A., Gambrell, J. L., & Stebbins, M. S. (2013). Minority performance on the Naglieri nonverbal ability test, second edition, versus the cognitive abilities test, form 6: One gifted program's experience. *Gifted Child Quarterly, 57*, 101–109. http://dx.doi.org/10.1177/0016986213477190

Hammill, D. D., Pearson, N., & Wiederholt, J. L. (2009). *CTONI-2: Comprehensive test of nonverbal intelligence* (2nd ed.). Austin, TX: Pro-Ed.

Harradine, C. C., Coleman, M. R. B., & Winn, D. C. (2014). Recognizing academic potential in students of color: Findings of U-STARS~PLUS. *Gifted Child Quarterly*, *58*, 24–34.

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

Jenkins, M. D. (1936). A socio–psychological study of Negro children of superior intelligence. *Journal of Negro Education*, *5*, 175–190. http://dx.doi.org/10.2307/2292155

Kornilov, S. A., Tan, M., Elliott, J. G., Sternberg, R. J., & Grigorenko, E. L. (2012). Gifted identification with Aurora: Widening the spotlight. *Journal of Psychoeducational Assessment*, *30*, 117–133. http://dx.doi.org/10.1177/0734282911428199

Krisel, S., & Cowan, R. (1997). Georgia's journey toward multiple-criteria identification of gifted students. *Roeper Review*, *20*(Suppl.), A1–A3.

Lohman, D. F. (2005a). Review of Naglieri and Ford (2003): Does the Naglieri nonverbal ability test identify equal proportions of high-scoring White, Black, and Hispanic students? *Gifted Child Quarterly*, *49*, 19–28. http://dx.doi.org/10.1177/001698620504900103

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

Lohman, D. F. (2009). Identifying academically talented students: Some general principles, two specific procedures. In L. V. Shavinina (Ed.), *International handbook on giftedness* (pp. 971–997). Berlin, Germany: Springer Science + Business Media.

Margolin, L. (1994). *Goodness personified: The emergence of gifted children*. New York, NY: Aldine de Gruyter.

Matthews, M. S. (2015). Creativity and leadership's role in gifted identification and programming in the USA: A pilot study. *Asia Pacific Education Review*, *16*, 247–256. http://dx.doi.org/10.1007/s12564-015-9373-x

Matthews, M. S., & Kirsch, L. (2011). Evaluating gifted identification practice: Aptitude testing and linguistically diverse learners. *Journal of Applied School Psychology, 27*, 155–180. http://dx.doi.org/10.1080/15377903.2011.565281

Matthews, M. S., Ritchotte, J. A., & Jolly, J. L. (2014). What's wrong with giftedness? Parents' perceptions of the gifted label. *International Studies in Sociology of Education*, *24*, 372–393. http://dx.doi.org/10.1080/09620214.2014.990225

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*, 258–278. http://dx.doi.org/10.1177/0016986216656256

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*, 69–89. http://dx.doi.org/10.1177/0016986213513794

McBee, M. T., Shaunessy, E., & Matthews, M. S. (2012). Policy matters: An analysis of district-level efforts to increase the identification of underrepresented learners. *Journal of Advanced Academics*, *23*, 326–344. http://dx.doi.org/10.1177/1932202X12463511

McCallum, R. S., Bell, S. M., Coles, J. T., Miller, K. C., Hopkins, M. B., & Hilton-Prillhart, A. (2013). A model for screening twice-exceptional students (gifted with learning disabilities) within a response to intervention paradigm. *Gifted Child Quarterly*, *57*, 209–222. http://dx.doi.org/10.1177/0016986213500070

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, *44*, 278–292. http://dx.doi.org/10.3102/0013189X15591157

Naglieri, J. A. (2008). *Naglieri nonverbal ability test, second edition* (NNAT-2). San Antonio, TX: NCS Pearson.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri nonverbal ability test (NNAT). *Gifted Child Quarterly*, *47*, 155–160. http://dx.doi.org/10.1177/001698620304700206

No Child Left Behind Act of 2001, 20 U.S.C. 70 § 6301 et seq. (2002).

Nomi, T. (2009). The effects of within-class ability grouping on academic achievement in early elementary years. *Journal of Research on Educational Effectiveness*, *3*, 56–92. http://dx.doi.org/10.1080/19345740903277601

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

Peters, S. J., Matthews, M. S., McBee, M. T., & McCoach, D. B. (2014). *Beyond gifted education: Designing and implementing advanced academic programs*. Waco, TX: Prufrock Press.

Peterson, J. S. (1999). Gifted—Through whose cultural lens? An application of the postpositivistic mode of inquiry. *Journal for the Education of the Gifted*, *22*, 354–383. http://dx.doi.org/10.1177/016235329902200403

Pfeiffer, S. I. (2012). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessment*, *30*, 3–9. http://dx.doi.org/10.1177/0734282911428192

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

Plucker, J., Hardesty, J., & Burroughs, N. (2013). *Talent on the sidelines: Excellence gaps and America's persistent talent underclass*. Storrs, CT: Center for Education Policy Analysis.

Plucker, J. A., Burroughs, N., & Song, R. (2010). *Mind the (other) gap! The growing excellence gap in K–12 education*. Bloomington, IN: Center for Evaluation and Education Policy.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. London, England: Psychology Press. (Original work published 1938)

Renzulli, J. S., & Gaesser, A. H. (2015). A multi criteria system for the identification of high achieving and creative/productive giftedness. *Revista de Educación*, *368*, 88–117.

Ross, P. O. (1993). *National excellence: A case for developing America's talent*. Washington, DC: U.S. Department of Education.

Siegle, D., Moore, M., Mann, R. L., & Wilson, H. E. (2010). Factors that influence in-service and preservice teachers' nominations of students for gifted and talented programs. *Journal for the Education of the Gifted*, *33*, 337–360.

Sweygers, A., Soetewey, K., Meeus, W., Struyf, E., & Pieters, B. (2009). Portfolios for prior learning assessment: Caught between diversity and

standardization. *Journal of Continuing Higher Education*, *57*, 92–103. http://dx.doi.org/10.1080/07377360902978194

Terman, L. M. (Ed.). (1925). *Genetic studies of genius* (vol. 1). Palo Alto, CA: Stanford University Press.

Vogl, K., & Preckel, F. (2014). Full-time ability grouping of gifted students: Impacts on social self-concept and school-related attitudes. *Gifted Child Quarterly*, *58*, 51–68. http://dx.doi.org/10.1177/0016986213513795

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*, 121–136. http://dx.doi.org/10.1177/0016986208330564