

Identifying Under-Served Student Populations for Gifted Programs: Some Methods and Frequently Asked Questions

The under-representation of low-income, second language learning, and culturally diverse students in gifted and talented programs is a continuing issue. It is important that all educators of the gifted be familiar with research supported, best practices in the use of assessments to identify these learners. This past winter, Dr. Scott J. Peters, an associate professor in the educational foundations department of the University of Wisconsin at Whitewater, came to Minnesota to share important considerations and research-based recommendations in identifying underrepresented learners for gifted education services. Below, he summarizes some of his recommendations as well as addresses some frequently asked questions.

Universal screening is always best. What this means is that whatever you use as the measure or measures of “giftedness” (ideally measures that predict success in the program with which students will be provided), it is always preferable to give those measures to all eligible students, as opposed to only those who meet some other initial criteria. Only giving them to a select number of students, such as following a referral or nomination by a teacher, is *always* less optimal from the standpoint of accuracy and sensitivity of identification. So, for example, testing all third graders for giftedness with a measure of general ability (e.g., the Cognitive Abilities Test – CogAT or Otis-Lennon School Ability Test - OLSAT) is preferable to only testing students who are referred by teachers or parents or those who scored at a certain level on a state achievement test. Non-universal screening for gifted identification will always result in more students being missed than if universal screening had been applied. What’s more, the negative effect of non-universal screening will disproportionately affect students from traditionally underrepresented populations, thereby exacerbating inequality in identification. See the paper listed below by Card and Giuliano on universal screening as well as McBee, Peters, and Miller (2016).

There is some middle ground between only testing those who are first nominated by a teacher and universally screening everyone. When a school only evaluates those who are first nominated, a relatively small percentage of students will end up being tested (probably less than 10 percent). Alternatively, under universal screening, 100 percent of students are evaluated for eligibility. A balance can be struck between these two to try and maximize benefits while also minimizing costs. For example, instead of only testing those who are nominated as “gifted” a school could instead ask teachers which students are at least “above average.” This would translate to evaluating closer to 50 percent of students for gifted

eligibility. Compared to universal screening, a school would only have to spend half the cost, but would still get almost all of the accuracy out of the identification system. For a more detailed explanation of this, see the McBee, Peters, and Miller (2016) article on my faculty website (listed above).

The use of high cut scores (e.g., 98th percentile or higher) is rarely defensible, and they will exacerbate underrepresentation. Even moving cut scores lower may still result in underrepresentation, but it will be less. Although lowering cut scores alone is not a solution to underrepresentation, overly high cut scores are just not logical in most cases and the existing instruments commonly used for identification don't have strong enough reliability at high score levels to tolerate the use of high cut scores. In other words, there is more error present at higher scores for pretty much any assessment instrument*. High cut scores might be attractive from the standpoint of smaller identified populations, but they will also result in many missed kids. As with non-universal screening, this negative effect will be felt disproportionately by students from traditionally underrepresented populations. See McBee, Peters, and Waterman (2014) and Peters and Gentry (2012) for a discussion of the effect of various cut scores on the size and diversity of the identified populations. The higher the cut scores, the smaller and less racially, ethnically, and socioeconomically diverse the identified population will be. This is due in large part to societal inequality of educational opportunity.

*This applies to any fixed-form, non-adaptive assessment such as most state achievement tests. Tests that are given out of level or that can adapt to student performance (computer adaptive tests such as the Measure of Academic Progress) do not suffer from this same reliability issue and will yield more-valid information at higher cut scores.

Using national norms for identification rarely make sense. Unless your gifted programs and interventions serve and draw students from a national or international audience (such as with Talent Search programs or many large universities), there is no defensible reason to make identification decisions on the basis of national or international norms. Instead, a student should be compared to other students within a particular educational context (such as a single school) in which his or her needs will be met. This means using local norms for identification where a student's performance is only compared to his or her peers at the same school. This is easy to do but it does make for identified students who have a wider range of learning needs, which will need to be taken into consideration when designing services. A "one-size-fits-all" approach to gifted services may not be appropriate when you have students with different levels of preparation. See the Lohman (2006) monograph, the Lohman and Renzulli (2007) paper, and the Lohman (2009) chapter for a discussion of local norms and when they are appropriate. Also see the [Lohman Excel spreadsheet](#) on how to create and use local norms. They might sound complicated and fancy, but really they're just a matter of sorting students test scores from highest to lowest and then applying your chosen "gifted" criteria.

Group-specific norms will reduce underrepresentation rates, but this comes at a cost. Instead of just comparing a student to others within his or her school (local norms), comparing a low-income student solely to other low-income students will serve to drastically reduce underrepresentation rates among low-income students (see Peters and Gentry, 2012). There have been a few papers that have shown this pretty conclusively. However, it also means that within a single school with a socioeconomically diverse population, students maybe identified

based on different cut scores and/or criteria. This has political and logistical challenges and implications for designing programs and supports that will benefit all identified students but if a school really wants to identify low-income students of potential, it makes sense. In most cases, a school or district would need to “compute” its own income- or language-group specific norms. Luckily, as with local norms, all this entails is a sorting of student scores by group. Once you have your data in an Excel spreadsheet, you need only sort those scores from highest to lowest, doing so for both low-income groups and higher-income groups separately. If the criteria for identification are scoring in the top 10 percent of each group and there are 50 low-income and 50 high-income students, then the top five in each group would be identified. Because educators don’t often have access to student income data, this is usually something that will need to be done at the district level.

Because racial and ethnic minority students are also much more likely to be from low-income families, using income-group specific norms will also identify larger numbers of racial / ethnic minorities. See Peters, Matthews, McBee, and McCoach (2014), Peters and Engerrand (2016), and Peters and Gentry (2012) as well as the Card and Giuliano study which applied this in one of the largest districts in the nation. When applied, group specific norms will identify larger number of underrepresented groups, but it is then up to the school to decide what services those students need in order to be successful. In this sense identification is the easy part.

Nonverbal ability tests will not solve the problem of underrepresentation – at least they haven’t so far. To call a test “nonverbal” is most often a reference to the fact that such tests require little to no language mastery in order to complete. Most often they involve figural reasoning tasks in which examples are presented and the student is able to understand the task expected of him or her without any need for written or spoken instructions. These tests show obvious appeal – they require no English language mastery, so it stands to reason they should show smaller observed test score differences across student subgroups, thus allowing for a solution to underrepresentation. It’s a nice idea, but it hasn’t panned out in research. It is safe to say the perpetual search for the “perfect test” that will eliminate underrepresentation is unrealistic and is actually a distraction from the larger issues of societal and educational inequality. In addition, research shows that nonverbal tests are not better at identifying typically under-represented gifted students than traditional tests of achievement or ability and have lower predictive validity for school performance. These tests may be good universal screeners for young children in the early grades, but so far they haven’t lived up to their promise of being culture-neutral measures of talent. See Lakin and Lohman (2011) as well as Peters and Engerrand (2016) for a discussion of nonverbal ability tests. The Peters and Engerrand paper also presents a few studies that have been done that used nonverbal tests to ID GT students. In short, even if nonverbal abilities tests were still valid measures of student need for advanced services, so far they have failed to identify larger numbers of underrepresented students than have more-traditional measures, in part because they still show large average score differences across student subgroups.

“Good identification” should always be measured by the degree to which the identification system locates students in need of services and who will be successful in a particular domain-specific program or intervention. General, domain universal gifted identification is much harder to do and the creation of an identification system without knowledge of what the

program will be impossible to do well. The design of the program has to precede the design of the selection criteria.

Structured performance and observation protocols have also shown some potential for better locating underrepresented learners. By “structured observation protocols,” I mean identification procedures that involve observing students with some kind of structured rubric as they engage in a task relevant to the program that would be provided to those identified by the task. In such protocols, individuals are given a task to complete and then trained observers watch for certain behaviors. See the 2015 article by Carol Horn on Young Scholars for a structured performance assessment method used to locate underrepresented learners for gifted programs. This program essentially uses challenging curriculum to identify students with gifted achievement potential. Another similar approach with older students is the use of problem based learning units in the article by Gallagher and Gallagher. Such systems of ID have shown promise, although they are logistically work intensive to implement. They also face a daunting task in overcoming the natural unreliability that comes with human raters. No matter how much training a group of humans receives, they are still humans and their perceptions are still influenced by their unique prior life experiences. This is why using individual teachers’ ratings as a *required* component in gifted student identification is not recommended.

[Dr. Peters Faculty Website](#)

Frequently Asked Questions

Won't universal screening be expensive?

Yes. It will be more expensive than the alternative (e.g., only testing those who are first nominated), but it will also miss far, far fewer students. It's a matter of balance. Testing zero percent of students will come at no cost but miss 100 percent of students. Universal screening will miss the fewest students but at the greatest cost. This is why we believe there is a middle ground as described above.

We can't lower cut scores because too many kids will be identified.

This is a fair concern, but there aren't a lot of other options. A way to combat this is to only apply the lower cut score to the underrepresented group (i.e., group specific norms), but that can add political and parent communication challenges. Put simply, finding and serving larger numbers of underrepresented students in gifted programs is either going to take additional resources or involve no longer serving some students from dominant cultural groups.

If we use local norms, what happens when kids move to a different school in the district? Or a different district?

The rationale for local norms is that they tell you which students are so advanced that they are likely to be going under challenged in their current instructional environment. If that instructional environment changes, such as when a student moves to an overall higher achieving school, she might now be able to be challenged in the "regular" classroom. It's not about which building you attend that decides if you are appropriately challenged. Regardless of the building, everyone should be appropriately challenged. The real question is just whether or not you can get challenged in the general, grade level classroom or if you need to receive a gifted service. Yes, a student with a score of 120 might be identified as gifted in one building but not another under local norms, but in theory this is fine because it means at the new building, she is effectively challenged outside of gifted services. This is a different way of thinking about "gifted" education, but it makes much more instructional sense. After all, the label isn't what matters. What matters is that all kids are appropriately challenged.

How do we "get" local norms?

Dave Lohman, a now-retired professor from the University of Iowa, has an excellent handout and example spreadsheet on his website on how to do this. It takes surprisingly little time. Some test companies will also provide these if asked (e.g., CogAT). If you have to do them on your own, it's as simple as sorting your student scores from highest to lowest and then taking the top X percent based on your chosen "gifted" criteria.

But aren't nonverbal ability tests culturally neutral and therefore better for gifted identification?

No and no. First, nonverbal ability tests still show large group differences meaning they are not any more “culturally neutral” in their resulting scores than many other academic assessments. Second, even if they were culturally neutral in their actual cognitive requirements, that doesn't make them good. Admitting students to college based on the first letter of their home town would be culturally neutral, but it also wouldn't yield the students who are most likely to benefit from college. This is a validity problem that plagues all “nonverbal” ability tests. The content they measure simply isn't related to the content of most gifted education programs. A district could make a program that was properly aligned to the skills and dispositions measured by nonverbal tests, but this isn't something I see very often.

Are group specific norms permissible?

Of course, this is something you would want to check with your local school district attorneys about since every state is different, but as long as they are not used based on ethnic or racial groups, then there shouldn't be an issue. Typically, this means language or income-group specific norms. See Peters and Engerrand (2016) for a detailed overview of this issue.

Why is underrepresentation such a pervasive issue?

The simple answer to a not at all simple question? In the absence of equal opportunities and access for all students, inequality will manifest itself in any measure of academic achievement or ability. It's not the fault of the tests that some student groups score lower. It's a result of society providing very different learning experiences to children based on where they live, who they are, and what they can afford. If the differences were due to flawed tests, this would be an easy fix. But they're not (for the most part). Instead, they're due to large differences in educational opportunity that show up any time students are assessed on academic content. This is also why multiple opportunities and pathways to be identified are important. If identification only happens once and via one pathway, students who have yet to develop the skills or talents measured by that process will be missed, thereby exacerbating underrepresentation.

References

- Card, D. & Giuliano, L. (2015). *Can universal screening increase the representation of low income and minority students in gifted education?* (National Bureau of Economic Research Working Paper No. 21519). Retrieved from [NBER website](http://www.nber.org/papers/w21519): (<http://www.nber.org/papers/w21519>).
- Gallagher, S. A., & Gallagher, J. J. (2013). Using problem-based learning to explore unseen academic potential. *Interdisciplinary Journal of Problem-Based Learning*, 7(1), 111-131.
- 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.
- 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, 595-623. doi: 10.1177/016235321103400404
- 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.
- Lohman, D. F., & Renzulli, J. (2007). A simple procedure for combining ability test scores, achievement test scores, and teacher ratings to identify academically talented children. Unpublished paper retrieved July 14, 2011 from the [Lohman Excel spreadsheet](#)
- Lohman, D. F. (2006). *Identifying academically talented minority students* (Research Monograph RM05216). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.
- 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.
- McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rules. *Gifted Child Quarterly*, 58, 69-89.
- Peters, S. J., & Gentry, M. (2012). Group-specific norms and teacher rating scales: Implications for underrepresentation. *Journal of Advanced Academics*, 23, 125-144.
- Peters, S. J., Matthews, M., McCoach, D. B., McBee, M. (2014). *Beyond gifted education: Designing and implementing advanced academic programs*. Waco, TX: Prufrock 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, 159-171.

Example Assessments by Type

Assessment Type	Example
Academic Achievement Test ¹	Measures of Academic Progress (MAP); Peabody Individual Achievement Test (PIAT); Minnesota Comprehensive Assessments (MCA); Iowa Test of Basic Skills (ITBS)
Nonverbal Ability Tests	Naglieri Nonverbal Ability Test; Test of Nonverbal Intelligence (TONI); Universal Nonverbal Intelligence Test (UNIT)
Academic Ability/Aptitude Tests ²	Woodcock-Johnson Tests of Cognitive Abilities; Slosson Intelligence Test; Stanford-Binet Intelligence Scales (SB-5); Weschler Intelligence Scale for Children (WISC)
Teacher Rating Scales	Scales for Identifying Gifted Students (SIGS); Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS); Gifted Rating Scales (GRS); HOPE Teacher Rating Scale

¹Although these are all measures of academic achievement, they differ in the underlying content standards that they measure. Some might measure the Common Core State Standards whereas others measure state-specific standards.

²These include both individually-administered and group-administered assessments. Some of these are tests of academic ability while still others measure intelligence or aptitude. Although often grouped together, there is disagreement over whether or not these assessments all measure the same thing.