Comparing Tests Used To Identify Ethnically Diverse Gifted Children:

A Critical Response to Lewis, DeCamp-Fritson, Ramage, McFarland, & Archwamety

Russell T. Warne

Recently *Multicultural Education* published a research article examining the effectiveness of three different tests in identifying ethnically diverse gifted and talented children (Lewis, DeCamp-Fritson, Ramage, McFarland, & Archwamety, 2007, Volume 15, Number 1, pp. 38-42). After examining the proportions of Caucasian and non-Caucasian children identified as gifted by each test, Lewis and her colleagues concluded, ". . . the Raven's Standard Progressive Matrices was a more effective means of selecting for ethnically diverse children who may be gifted than one example of a traditional achievement test [the Iowa Tests of Basic Skills] or even the newer Naglieri Nonverbal Abilities Test" (p. 42).

It is my opinion that those researchers' conclusion, however, must be questioned because of an assortment of philosophical, testing, and statistical issues that they did not consider when conducting their study. The purpose of this response is to use the Lewis et al. article as a springboard to discuss the issues in gifted education that Lewis and her colleagues overlooked and to consider how those issues relate to multicultural education.

I wish to make it clear at the onset of this article that I do not disagree with the goals of Lewis et al.'s (2007) study. The field of gifted and talented education nearly universally recognizes the need to identify more ethnically diverse and economically disadvantaged students, who are typically underrepresented in programs that serve the gifted and talented (Abell & Lennex, 1999; Hunsaker, 1994; McBee, 2006) and leaders in the field strongly urge more research on culturally diverse gifted students (e.g., VanTassel-Baska, 2006). In other words, I wholeheartedly agree that something must be done to improve the identification of gifted and talented minority students.

Russell T. Warne is a doctoral student in the Research, Measurement, and Statistics Program of the Department of Educational Psychology at Texas A&M University, College Station, Texas.

A Note from the Editors

This article is a response to a previous research article published in *Multicultural Education* in 2007. We include the response in this issue to continue the dialogue initiated in the original article.

We encourage readers of *Multicultural Education* to submit responses to articles published in the magazine for consideration.

This article will cover three major areas of discussion regarding Lewis et al.'s (2007) study. The first section will cover theoretical issues related to giftedness. The second section will discuss the tests that were used in the study, while the third section will discuss the statistics of Lewis and her colleagues' study. The article will then conclude with a discussion of the latest trends in gifted education for finding and serving ethnically diverse students.

Theoretical Issues

What is giftedness? Like most definitions of a psychological construct, there is not unanimous agreement among professionals. Lewis and her colleagues (2007) utilized the federal government definition, which uses the term "giftedness" to refer to

> . . . students, children, or youth who give evidence of high achievement capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who need services or activities not ordinarily provided by school in order to fully develop those capabilities. (Pub. L. No. 107-100, Title IX)

Although this legal definition is helpful, it only tells us that giftedness can come in many domains. The federal government definition tells us nothing of the origin or nature of giftedness.

The best definition of giftedness that has emerged from the research comes from the Columbus Group, a group of researchers, practitioners, and parents who met in Columbus, Ohio, in 1991. According to the Columbus Group,

> Giftedness is asynchronous development in which advanced cognitive abilities and heightened intensity combine to create inner experiences and awareness that are qualitatively different from the norm. This asynchrony increases with higher intellectual capacity. The uniqueness of the gifted renders them particularly vulnerable and requires modifications in parenting, teaching, and counseling in order for them to develop optimally. (Morelock, 1992, p. 14)

In other words, giftedness is a case of accelerated development that results in inherently different psychological experiences than what is typical for a person's age and experience.

Therefore, giftedness is not defined by test scores alone. Giftedness is more like a cluster of abilities and behaviors that tend to be comorbid with one another. Some of these behaviors are intellectual (Konstantopoulos, Modi, & Hedges, 2001; Lubinski, 2000; Ruf, 2005), and some are not (Brink, 1982; Daniels & Piechowski, 2009; Rimm, 1984; Towers, 1987; Speirs Neumeister, Williams, & Cross, 2007; Webb, Amend, Webb, Goerss, Beljan, & Olenchak, 2005). However, no gifted child displays all of these behaviors and it would be inaccurate to talk about "the typical gifted person," because gifted children (and adults) are a diverse group (McGuffog, Feiring, & Lewis, 1987; Robinson, 1981).

In their 2007 article, Lewis and her colleagues didn't emphasize the richness of giftedness. Instead, she and her coauthors defined giftedness as scoring at or above the 80th percentile on one of the tests they

examined. Moreover, they did not examine whether the children who score highly on the three tests—the Raven's Progressive Matrices (Raven's; Raven, 1947), the Naglieri Nonverbal Abilities Test (NNAT; Naglieri, 1997), and the Iowa Tests of Basic Skills (ITBS; Hoover, Dunbar, & Frisbie, 2001)—were actually gifted or not.

Another apparent shortcoming of Lewis et al.'s study was that the authors didn't say whether they were looking for children who displayed any particular sort of giftedness, or whether they were searching for children who were gifted in certain areas. For example, if one uses the federal government's definition of giftedness, then children may be gifted in leadership or creative endeavors (among other domains). It is doubtful that any of the three tests from the Lewis et al. study is useful for identifying children of any ethnicity who display their giftedness in one of these areas.

Another important theoretical problem in the original Lewis et al. (2007) article is that the authors treat intelligence as a psychological construct that can be separated from experience and culture. Even though IQ scores are remarkably stable across a person's lifetime (Moffitt, Caspi, Harkness, & Silva, 1993), intelligence is not divorced from other psychological processes and influences, including culture.

It has become very popular in recent years to claim that nonverbal tests like the Raven's and NNAT are fairer for culturally diverse students than verbal tests, such as the ITBS. Proponents of nonverbal intelligence testing claim that verbally-based tests are culturally biased against diverse test takers and that the tests should not be used to identify diverse gifted students (e.g., Naglieri & Ford, 2003). Lewis et al. agreed with this popular notion in their literature review, saying, ". . . alternative methods of selection that are not based on acquired academic skills *or verbal abilities* may be necessary [to identify minority gifted children]" (2007, p. 38, emphasis added).

There are many problems with this viewpoint. First, culture is not a patina that must be stripped away in order to examine the real underlying construct of interest. Rather, culture is a rich, multidimensional sociopsychological quality that is completely inseparable from every thought, action, and emotion that a person experiences. No test, no matter how well-designed, can temporarily remove any portion of somebody's culture from their mind in order to measure a construct (Lohman, 2006a). Moreover, to attempt to do so may be a disservice to the rich and varied cultures that test takers come from.

Another problem associated with assuming that nonverbal tests of intelligence are fairer than verbal tests is that the terms "verbal" and "nonverbal" only refer to the *stimuli* used in a test, not the psychological processes needed to solve test questions (Lohman, 2005). Indeed, questions presented in a verbal format may require spatial reasoning to answer, and nonverbal test items may require verbal mental skills to respond to. Paradoxically, it is theoretically possible to have a so-called nonverbal test that requires verbal skills to answer every question.

Finally, proponents of nonverbal testing often do not realize that by examining intelligence only through nonverbal means, the risk of construct underrepresentation increases. *Construct underrepresentation* occurs when a test fails to investigate every facet of a construct and thereby distorts the understanding of the construct. Although intelligence has been boiled down to one high-order *g* factor for more than a century (Spearman, 1904), theorists agree that intelligence has two major facets—a verbal component and a nonverbal component (Horn & Cattell, 1966). By only measuring one of these major facets, a nonverbal test only presents half of the picture of someone's intellectual ability. Moreover, by not measuring a construct completely, a test is more likely to introduce error, which reduces the reliability of obtained scores and makes any judgments based on those scores more inaccurate (Lohman, 2005).

Test Issues

Lewis et al. (2007) compared results of three different tests in order to determine which one was best at identifying ethnically diverse gifted children. Each test will be discussed alone before they are compared and issues surrounding their use by Lewis and her colleagues are explored.

Raven's Progressive Matrices

The Raven's purports to measure nonverbal intelligence through the use of nonverbal stimuli. The test is probably the oldest nonverbal test in use and is still very popular today. However, leading psychometricians advise against using the Raven's for many reasons. First, the norms for the Raven's are highly distorted and the test has never been normed correctly for an American population (Lohman, Korb, & Lakin, 2008; Mills & Ablard, 1993), a fact that Lewis et al. (2007) recognized. Unfortunately, inaccurate or distorted norms mean that *every comparison* using those norms is also distorted. Therefore, any judgment (such as labeling a child as gifted or not gifted) made on the basis of a Raven's score carries a much greater risk of being inaccurate.

The Raven's also has the problem of the Flynn effect (Flynn, 1987), in which scores on intelligence tests gradually inflate over time. If norms are not adjusted periodically, most intelligence tests will inflate scores three points after a decade with the same norms. The Raven's, however, inflates scores by an average of eight points (.5 *SD*) per decade—nearly three times the rate for a verbal IQ test (Flynn, 1987, p. 186). Therefore, even if the Raven's had accurate norms, they would be out of date much more quickly than those of a verbal intelligence test. The Raven's norms were last updated in 1998 (Kaplan & Saccuzzo, 2005, p. 342).

Naglieri Nonverbal Abilities Test

According to its creator, the NNAT ". . . is a brief nonverbal measure of ability that does not require the child to read, write, or speak" (Naglieri & Ford, 2003, p. 157). Modeled after the Raven's, while attempting to correct for the psychometric deficiencies associated with the earlier test, the NNAT is growing in popularity among school psychologists.

The NNAT has the same problem as the Raven's with the Flynn effect, due to the NNAT's nature as a nonverbal measure of intelligence. In 2007, when Lewis et al. were writing, the NNAT norms were a decade old. However, they have since been updated with the revision of the NNAT that was published in 2008, which means the new norms are not yet outdated enough to exhibit the Flynn effect.

Iowa Tests of Basic Skills

Unlike the Raven's and NNAT, the ITBS is an achievement test designed to measure ". . . growth in fundamental areas of school achievement . . ." (Hoover, Dunbar, Frisbie, Oberley, Ordman, Naylor et al, 2003, p. 1). The ITBS is one of the oldest and most venerable academic achievement tests on the market and has over six decades of strong research and psychometrics behind the latest editions. The ITBS consists of nine vertically aligned levels from kindergarten to grade 8 and has a sister test—the Iowa Tests of Educational Development (Forsyth, Ansley, Feldt, &

Alnot, 2001)—that continues the ITBS's scale through the 12th grade.

The ITBS is not subject to the Flynn effect, due to its status as an achievement test. Indeed, achievement levels have remained stable, in comparison to IQ scores, since the 1950s (Hoover et al., 2003, p. 57-59).

Tests Compared

Because the ITBS measures academic achievement and the NNAT and Raven's claim to measure nonverbal intelligence, it is not fair to compare these instruments apples to apples in the way that Lewis et al. (2007) did. To do so would be like comparing one athlete's time on the 100 meter dash to another athlete's time on a downhill skiing course. In fact, Lewis et al.'s Table 2 (p. 41) shows that none of the three tests correlates with any other at a level higher than r=+.52. This means that no more than 27% of the variance of one set of test scores can be accounted for by the variation in scores on another test. Clearly, these three tests do not measure the same construct, even though the creators of the Raven's and the NNAT claim that those two tests do.

Because each of the three tests measures a different psychological construct in Lewis et al.'s (2007) sample, one should *not* expect to identify the same proportions of children. Indeed, if we accept Lewis et al.'s definition of giftedness as scoring at or above the 80th percentile on any of the tests, the Raven's identified over double the number of children that either of the other tests did. Yet, Lewis and her coauthors still assumed that all three tests measured intelligence equally well and based their article's final conclusions on that assumption, even though the data did not support it.

Knowing that each test was examining a different construct tells us nothing about which test—if any—examines intelligence. To do so, we must examine how each test correlates with generally accepted measures of intelligence. Unfortunately, Lewis et al. (2007) did not perform such a comparison. Instead, they chose to accept the test creators' claims of what their tests measured and didn't undertake to examine the technical manuals of either the NNAT or the ITBS. To investigate which tests were examining which constructs, Lewis and her coauthors could have calculated correlations between the test scores and manifestations of intelligence, such as grades, individually administered IQ tests, or performance on problem solving tasks.

Fortunately, other researchers have performed such correlations in validity studies. Mills and Tissot (1995) correlated Raven's scores against school grades in math and English. Given the "nonverbal" nature of the Raven's, it is surprising to learn that the Raven's was a better predictor of English grades (r=+.20) than math grades (r=+.17); the correlation with math grades was not statistically significant, although the correlation with English grades was (p. 214). The Raven's is a notoriously poor predictor of academic performance, although it does have moderate correlations (between +.40 and +.65) with other tests of reasoning abilities (Mills & Ablard, 1993). These moderate correlations with other tests of reasoning abilities are good evidence that the Raven's does test some sort of fluid reasoning but is likely still subject to construct underrepresentation. The low predictive power for academic success, however, should be troubling to those who plan to use the Raven's tests for admission to highly demanding academic programs, such as gifted and talented programs.

The NNAT performs at least as well as the Raven's on comparisons with nonverbal intelligence and achievement. Non-English language learners' NNAT scores correlated moderately with their scores on the Cognitive Abilities Test (CogAT; Lohman & Hagen, 2001), which measures fluid reasoning. NNAT scores correlated r=+.44, +.55, and +.66 on the Verbal, Quantiative, and Nonverbal portions of the CogAT respectively for non-English language learners (Lohman et al., 2008, p. 289). Although this is an improvement, the CogAT and the NNAT do not measure intelligence equally well, as can be seen by the fact that the explained variance ranges from 19% to 44%. This is most likely a reflection of the construct underrepresentation inherent in all nonverbal intelligence tests.

Lohman et al. (2008) also compared NNAT scores with reading and mathematics composite scores on the TerraNova Test, which examines academic achievement. The correlations were r=+.35 and +.55, respectively for English language learners (p. 290). For non-English language learners, the correlations were similar: r=+.38 and +.48. Lohman et al. also reported additional correlations for the NNAT for grades 3, 4, and 5-6, with correlations ranging between r=+.16 and +.66. Although this is an improvement over the Raven's performance, it still means that between 4% and 44% of the variance in a child's reading or mathematics ability can be explained through their NNAT score. If the NNAT were a good measure of intelligence, all these correlations would be higher (at least +.70, which would allow both variables to share at least half their variance) and more consistent across groups.

Of the three tests examined in Lewis et al.'s (2007) study, the ITBS has the most data about its ability to measure its intended construct. The ITBS, of course, is not designed as an intelligence test, but rather an achievement test. Therefore, ITBS scores should correlate highly with other measures of achievement, which is exactly what has been observed. Grade 6 and grade 8 ITBS scores correlate +.73 and +.78 with scores on the American College Test, which is a popular test for college admission (Hoover et al., 2003, p. 47). Also, grade 8 ITBS scores correlate well with high school GPA, between +.38 and +.61 (Hoover et al., 2003, p. 47). These high correlations show that the ITBS examines exactly what its creators claim it measures: academic achievement.

Nevertheless, the ITBS is also an adequate measure of intelligence and may suffice when large scale individual testing is impractical (such as in a large school). Indeed, ITBS scores from childhood correlated +.64 with study participants' midlife IQ on the Wechsler Adult Intelligence Scale (Spinks, Arndt, Caspers, Yucuis, McKirgan, Pfalzgraf et al., 2007, p. 565). Apparently, childhood ITBS scores are about as good of a measure of adult intelligence as childhood NNAT scores are of current intelligence.

In sum, none of the three tests that Lewis et al. (2007) used in their study is a high quality measure of intelligence, and only the ITBS measures its intended construct (academic achievement) well. One should keep this fact in mind before using *any* of these tests for identification for gifted and talented programs. I suspect that because the tests measure different constructs, different children scored highly on each test and that no two tests would have selected the same children for admission into a gifted program. Unfortunately, Lewis and her coauthors did not report any data that could support or disprove this supposition.

Statistical Issues

There are also some statistical issues that Lewis et al. (2007) did not examine when conducting their study. These are statistical issues that all quantitative substantive researchers must grapple with, not just those in gifted education.

First, Lewis and her coauthors did

not report the reliability of the scores that their sample obtained on the tests, even though this is widely recognized as best practice (Warne, 2008; Wilkinson & Task Force on Statistical Inference, 1999). Although Lewis et al. did report the reliability of the samples used to norm each test (p. 40), this is not sufficient because *a reliability coefficient pertains only to the sample from which it was derived and not to any other sample* (Thompson & Vacha-Haase, 2000). To take a reliability coefficient from a sample and apply it to another sample is called *reliability induction* and is an inherently erroneous practice (Shields & Caruso, 2004; Vacha-Haase, Henson, & Caruso, 2002). Low reliability attenuates correlations and makes decisions based on obtained scores less accurate and consistent. Because they did not report their reliability coefficients, it is impossible for the reader of Lewis et al.'s article to know whether the different proportions of each ethnicity that each test identified as gifted in the study is due to the different nature of the tests or low score reliability.

Another statistical problem that the article presented is the use of percentile scores in their statistical analysis. Percentiles, while useful in comparing scores across tests, are ordinal data, which makes taking the mean and standard deviation of percentiles meaningless (Thompson, 2006). It is also not meaningful to use percentiles as the dependent variable in an analysis of variance (ANOVA), as Lewis and her colleagues did. Percentile scores have these limits because the distances between adjacent scores are not equal; the further from the median scores are, the larger the difference between adjacent scores grows. Lewis and her coauthors should have instead used standardized scores on each test's own metric.

This property of percentiles makes studying gifted children particularly difficult, because they appear more alike than they really are. For example, if two first-graders are in the 99th percentile for math ability, one of them may be able to do simple multiplication and the other one may be able to do pre-algebra. Even though their percentile score is the same, their mathematical abilities are different. This is a fact that all researchers in gifted education must deal with.

Discussion on Ethnically Diverse Students and Gifted Education

My intention in this article is to use the Lewis et al. (2007) study as a lens through which we can examine gifted and talented education from a multicultural perspective. I have not pointed out their study's pitfalls and shortcomings merely to criticize their work. The problems I have pointed out in their article are not uncommon in gifted education and other quantitative educational research. What I wish to stress is my agreement with the overall intention of their study—to examine ways to serve more ethnically diverse gifted and talented students.

What are we to do about diverse gifted students? The answer is not to rely on nonverbal measures of intelligence to increase those students' enrollment in gifted and talented programs. Gifted programs are usually highly verbal and rely on large mental reservoirs of information. To determine admission into such a program with a test that ignores these qualities is a surefire formula for choosing students who are a poor fit for the program (Mills & Ablard, 1993).

Moreover, such a procedure is far more likely to overlook students of all ethnicities who would flourish in the program. The greatest predictor of future academic success is current academic success and the second strongest predictor is verbal ability (Lohman, 2006b). This is true for all ethnic groups and all levels of English mastery (Lohman, 2005). Furthermore, relying exclusively on nonverbal tests of ability provides a very limited view of a child's intellectual ability. Any admissions criteria into a gifted program *must* take these facts into account.

To compensate, though, for the low proportion of diverse students in gifted programs, several procedures have been suggested. One of the most successful treatments is called *front loading* (Mills et al., 1992; Nedler & Sebera, 1971; Briggs, Reis, & Sullivan, 2008). In front loading, administrators identify promising diverse and low-SES students who fall short of normal criteria for inclusion in gifted and talented programs. Those children are then funneled into intensive programs that build up their study skills, verbal ability, factual knowledge, and academic performance until the students do qualify for the main gifted program, usually after at least a year of intervention. In other words, front loading "primes the pump" of academic ability and allows ethnically diverse students to prepare for the academic challenge of the full gifted program in a safe environment.

Front loading not only helps gifted diverse students qualify for gifted programs, but also lessens the high attrition rate of ethnically diverse students from gifted education programs. Indeed, for some researchers the main problem with diversity in gifted education isn't identification. Rather, the problem is retaining ethnically diverse students in gifted programs (e.g., Moore, Ford, & Milner, 2005). The preparation that front loading gives to these students may have a positive impact on the attrition rates of culturally diverse students in gifted programs, which will in the long run raise their representation in gifted education programs.

Gifted minorities also seem to prosper under mentorships and it is suggested that mentorships are a highly effective way to serve ethnically diverse students and retain them in gifted programs (Briggs et al, 2008). Gifted African American males, in particular, perform well under conditions that create meaningful relationships between the student and a fellow African American who has already achieved success academically or professionally (Bonner, 2003).

Authorities in gifted education also agree that creating connections between the school and the homes of ethnically diverse gifted students is essential (Briggs et al., 2008; Ford, 1998). One of the most robust findings in education is that students whose parents are involved with their education perform better in academics than students whose parents are not (Jeynes, 2003). Culturally diverse gifted students' parents are much less likely to be able navigate a school system's bureaucracy and may feel alienated by the unfamiliar terminology and procedures of a gifted program.

Recruiting and retaining qualified minorities into gifted programs therefore requires administrators to build bridges between the school and those students' homes. Such connections require more than just sending a note to the parents in a child's backpack; extra time and dedication on behalf of the school administrators are required, but such efforts will likely be rewarded (Briggs et al., 2008; Ford, 1998, 2003).

Finally, teachers of both gifted and regular students need additional training in spotting giftedness in all its manifestations in children of all backgrounds (Ford, 1998). Most teachers have little or no training in giftedness (Ruf, 2005), and teachers are less likely to recommend non-Asian minority students than Caucasian or Asian students for gifted programs (Ford, 2003; McBee, 2006). By raising teachers' awareness of the signs of giftedness, fewer ethnically diverse gifted students are

likely to languish in mainstream education classes.

Conclusion

In their study, Lewis and her colleagues (2007) brought a welcomed critical eye to the state of culturally diverse gifted students in the United States. By bringing these issues into the harsh light of scrutiny, educators, researchers, and administrators are more likely to become agents of positive social change in the lives of culturally diverse gifted students. As these students are better served by their schools, our nation can only benefit as these gifted children grow up into gifted adults and make valued contributions to society.

However, the Lewis et al. (2007) study and its approaches to testing do have some flaws. The Raven's Standard Progressive Matrices test is *not* a good instrument for identifying underserved gifted students (Mills & Ablard, 1993; Mills et al., 1995) and *any* nonverbal test in general is not a good identification tool for students who are likely to succeed in gifted programs (Lohman, 2005).

Instead, those who wish to identify, serve, and retain ethnically diverse gifted students in gifted programs have a wide variety of tools at their disposal to help this underserved population. The interested reader should consult Briggs et al. (2008) and Ford (1998, 2003) for suggestions that have strong data supporting their effectiveness. These methods are more difficult than merely "picking the right test," but they will produce more lasting positive benefits for ethnically diverse gifted students and the programs that serve them.

References

Abell, D. J., & Lennex, L. (1999). *Gifted education: Don't overlook the disadvantaged*. Point Clear, AL: Mid-South Educational Research Association. (ERIC Document Reproduction Service No. ED436918)

Bonner, F. A., II (2003). To be young, gifted, African-American, and male. *Gifted Child Today, 26*, 26-34.

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*, 131-145.

Brink, R. E. (1982). The gifted preschool child. *Pediatric Nursing, 9*, 299-303.

Daniels, S., & Piechowski, M. M. (Eds.). (2009). *Living with intensity*. Scottsdale, AZ: Great Potential Press.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

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

Ford, D. Y. (2003). Equity and excellence: Culturally diverse students in gifted education. In N. Colangelo & G. A. Davis (Eds.), *Handbook of Gifted Education* (pp. 506-520). Boston: Allyn & Bacon.

Forsyth, R. A., Ansley, T. N., Feldt, L. S., & Alnot, S. D. (2001). *Iowa tests of educational development*. Itasca, IL: Riverside Publishing.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa tests of basic skills, forms A, B, and C*. Itasca, IL: Riverside Publishing.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., et al. (2003). *Iowa Tests of Basic Skills guide to research and development*. Itasca, IL: Riverside Publishing.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57*, 253-270.

Hunsaker, S. L. (1994). Adjustments to traditional procedures for identifying underserved students: Successes and failures. *Exceptional Children, 61*, 72-76.

Jeynes, W. H. (2003). A meta-analysis: The effects of parental involvement on minority children's academic achievement. *Education & Urban Society, 35*, 202-218.

Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (6th ed.). Belmont, CA: Thomson Wadsworth.

Konstantopoulous, S., Modi, M., & Hedges, L. V. (2001). Who are America's gifted? *Journal of Education, 109*, 344-382.

Lewis, J. D., DeCamp-Fritson, S.S., Ramage, J. C., McFarland, M. A., & Archwamety, T. (2007). Selecting for ethnically diverse children who may be gifted using Raven's Standard Progressive Matrices and Naglieri Nonverbal Abilities Test. *Multicultural Education, 15*(1), 38-42.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*, 111-138.

Lohman, D. F. (2006a). Beliefs about differences between ability and accomplishment: From folk theories to cognitive science. *Roeper Review, 29*(1), 32-40.

Lohman, D. F. (2006b, May 22). Identifying academically gifted children in a linguistically and culturally diverse society. Invited presentation at the Eighth Biennial Henry B. & Jocelyn Wallace National Research Symposium on Talent Development. Retrieved 10 December 2008, from http://faculty.education.uiowa.edu/dlohman/pdf/LohmanWallace%202006%20talk.pdf

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive abilities test (Form 6)*. Itasca, IL: Riverside Publishing.

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*, 275-296.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology, 51*, 405-444.

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education, 17*, 103-111.

McGuffog, C., Feiring, C., & Lewis, M. (1987). The diverse profile of the extremely gifted child. *Roeper Review, 10*(2), 82-89.

Mills, C. J., & Ablard, K. E. (1993). The Raven's Progressive Matrices: Its usefulness for identifying gifted/talented students. *Roeper Review, 15*(3), 183-186.

Mills, C. J., & Tissot, S. L. (1995). Identifying academic potential in students in under-represented populations: Is using the Ravens Progressive Matrices a good idea? *Gifted Child Quarterly, 39*, 209-217.

Moffitt, T. E., Caspi, A., Harkness, A. R., & Silva, P. A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology & Psychiatry & Allied Disciplines, 34*, 455-506.

Moore, J. L., III, Ford, D. Y., & Milner, H. R. (2005). Recruitment is not enough: Retaining African American students in gifted education. *Gifted Child Quarterly, 49*, 51-67.

Morelock, M. J. (1992). Giftedness: The view from within. *Understanding Our Gifted, 4*(3), 11-15.

Naglieri, J. A. (1997). *Naglieri nonverbal ability test*. San Antonio, TX: The Psychological Corporation.

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.

Nedler, S., & Severa, P. (1971). Intervention strategies for Spanish-speaking preschool children. *Child Development, 42*, 259-267.

Raven, J. C. (1947). *Coloured progressive matrices*. London: H. K. Lewis.

Rimm, S. (1984). The characteristics approach: Identification and beyond. *Gifted Child Quarterly, 28*, 181-187.

Ruf, D. L. (2005). *Losing our minds: Gifted children left behind*. Scottsdale, AZ: Great Potential Press.

Shields, A. L., & Caruso, J. C. (2004). A reliability induction and reliability generalization study of the Cage Questionnaire. *Educational and Psychological Measurement, 64*, 254-270.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15* (201-293).

Speirs Neumeister, K. L., Williams, K. K., & Cross, T. L., (2007). Perfectionism in gifted high-school students: Responses to academic challenge. *Roeper Review, 29*(5), 11-18.

Thompson, B. (2006). *Foundations of behavioral statistics*. New York: The Guildford Press.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174-195.

Towers, G. M. (1987, April). The outsiders. *Gift of Fire, 22*. Retrieved 6 December 2008, from http://www.prometheussociety.org/articles/Outsiders.html

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*, 562-569.

VanTassel-Baska, J. (2006). NAGC symposium: A report card on the state of research in the field of gifted education. *Gifted Child Quarterly, 50*, 339-341.

Warne, R. T. (2008). *Reliability generalization (RG) analysis: The test is not reliable*. New Orleans, LA: Southwest Educational Research Association. (ERIC Document Reproduction Service No. ED500497)

Webb, J. T., Amend, E. R., Webb, N. E., Goerss, J., Beljan, P., & Olenchak, F. R. (2005). *Misdiagnosis and dual diagnoses of gifted children and adults*. Scottsdale, AZ: Great Potential Press.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.