Donna Y. Ford, April J. Lisbon, and Natasha Little-Harrison

#Black Intellect Matters: Inequitable Practices Yield Inequitable Results

The authors question why nonverbal measures are not used more frequently when assessing minority groups.

Terrence's Story

Terrence was a five-year-old African-American kindergartener who grew up in a home with his mother, grandmother, and two younger siblings in an urban community comprised of mainly low-income families. Terrence's family required government assistance in order to subsidize his mother's part-time job at a local restaurant. Terrence's family worked with him daily to prepare him for kindergarten. They read to him, and taught him how to count from 1 to 10 as well as the alphabet. Terrence's mother and grandmother believed he was ready to enter kindergarten.

After the first week of school, Terrence's teacher expressed concerns about his academic readiness for kindergarten to the school administrator. She did not share the family's view about Terrence's academic skills and readiness during her discussion with the school administrator. Terrence's teacher, Ms. Johnson, a White female, expressed that he was significantly behind his peers, most of whom were White and higher income. His teacher shared that Terrence struggled to identify both upper and lowercase letters of the alphabet when presented on a chart. The only letters he recognized were those in his name. He was also unable to identify numbers that were visually presented to him. Ms. Johnson feared that if he did not get some assistance immediately, Terrence would repeat kindergarten. Although the school psychologist questioned the appropriateness of evaluating the student with limited intervention data, the majority of the school-based team agreed that Terrence should be evaluated for special education services while simultaneously collecting intervention data.

Terrence was given a variety of assessments by the school psychologist, an

African-American female, as a part of the referral to evaluate him for special education services. The battery of testing would be used to assist the school-based team in determining his strengths and weaknesses as a kindergarten student and his need for specialized instruction. The results of the battery of testing proved of interest as the results showed that when the school psychologist asked questions using language familiar to Terrence, he responded correctly, although credit was not given for his responses.

For Terrence, it was not that he was unable to process academic information or that he was significantly behind; Terrence knew the information just like his peers. Instead, it was the manner in which questions were posed that confused Terrence. Simply put, Terrence was unfamiliar with traditional academic language. He was used to the manner in which questions were posed within his natural environment. If the school psychologist had not probed further on the IQ measure, Terrence clearly would have been (mis)identified as a slow learner based on his full scale intellectual quotient of 75 on the Reynolds Intellectual Assessment Scales (RIAS). His verbal skills were in the mild intellectual disability range (SS: 63), while his performance skills were on the lower end of average (SS: 92). The Comprehensive Test of Nonverbal Intelligence Second Edition (C-TONI-2) was further administered and confirmed that Terrence indeed had average cognition (SS: 93).

Introduction

The use of assessment to address student achievement and need for additional assistance (e.g., specialized instruction) is at the forefront of education. This practice, particularly as its use increases, is marked by controversy. Testing and interpretation of the results of students of color has an extensive history of complaints, debates, and litigation regarding the appropriateness and fairness of standardized testing to determine special education and gifted education eligibility for students of color. Gould (1996), Herrnstein and Murray (1994), and the National Center for Fair and Open Testing are only three examples of the polemic and contentious debates. Herrnstein and Murray in particular have been 'credited' with evoking and reviving racist arguments about the intelligence of Black students, and advancing genetic and cultural inferiority explanations. Gould was less polemic and more detailed by presenting extensive historical stories of deliberate work promoted to advance genetic argument that Whites were more intelligent than Blacks and that environment was trivial.

Because it is impossible to conduct equitable assessments of culturally different students using a single instrument and/or methodology (Ortiz & Ochoa, 2005), suitability and fairness of standardized intelligence tests continue to be questioned. When fairness is considered, when the scores obtained by Black and White students (where Black students' scores are lower), an important interpretation must be that problems lie within the test rather than in student ability or aptitude.

Statistically, an error in value appraisal is referred to as bias (Reynolds & Suzuki, 2012), which may result in an over- or under-estimation of students' cognitive and academic abilities. Under-identification of students may deprive them of services (such as gifted education), while over-identification may result in inappropriate placements (often evident in special education) (Mercer & Pullen, 2009). The examination of potential test, and cultural bias and cultural loading must be considered when evaluating culturally different students. Scholars admit to staunchly divergent views of standardized cognitive testing and its efficiency in determining Black students' best or fairest measure of ability and, accordingly, the most appropriate levels and types of service provision. Few proponents of standardized intelligence tests believe the

tests to be the perfect measure of human personality and ability (Sattler, 2008); many suggest that testing provides a universal standard by which to measure learned cognitive skills across racial and ethnic groups (Reynolds & Suzuki, 2012).

Challengers and opponents of intelligence testing as the sole factor of special education and gifted eligibility postulate the comparison of abilities among racial and ethnic lines to be less related to one cultural group's academic prevalence and, contend that it is, instead contingent upon a student's time of assimilation into mainstream culture. Salvia and Ysseldyke (2004) maintained that cognitive test items may be considered biased if they rely on a student's cultural knowledge as a measure of aptitude, as "the test will necessarily be invalid because it will also test the student's knowledge of American culture" (p. 174). Because the most widely used assessment measures in the United States are largely reflective of the cultural patterns and values of mainstream society (White and middle or upper class), the probability is high that many Black students' experiences will be different than what is considered to be 'normal'. Use of these traditional cognitive measures inflates the likelihood of Black students being subjected to test biases (cultural and linguistic loadings) on

standardized tests, and they may obtain lower scores that mask their knowledge, skills, and potential. The noted difference in scores between Black and White students (about 15 points, or one standard deviation from the national mean) has led to the development of a cultural test bias hypothesis (Brown et al., 1999; Ford & Helms, 2012; Reynolds, 1982a, 1982b; Reynolds & Brown, 1984b; Reynolds & Suzuki, 2012) which suggests mean performance differences to be a reflection of the measurement process rather than variability among racial and ethnic groups (Ford, 2013).

When examining the potential bias of educational cognitive tests, it is important to address the social consequences of limited access to educational opportunities and resources. Relatedly, critics posit that criterion- and norm-referenced tests are not designed to assess school quality, including curriculum, delivery, and teacher expectation and perception—areas directly related to student learning and achievement (Kelleghan, Madaus, & Airasian, 1982). The dissimilarity in the allocation of educational capital has led to multiple litigation cases (Russo & Ford, 2015), many covered by *Frontline*, *PBS Newshour*, and the Center for American Progress (see Appendix A).

History of Mis-Testing

Standardized tests have become the most effective racist weapon ever devised to objectively degrade Black minds and legally exclude their bodies (Kendi, 2016)

Test bias and how individual measures of intelligence are dissected and contrived by theorists - new and old - continue to be a source of contention among critics and scholars alike. In *The Mismeasure of Man* (1996), Gould presented various historical cases studies which are relevant today that bogusly sought to degrade the intellectual abilities of individuals other than those of European descent. Gould exposed how various scholars used science, specifically biology, in order to promote the ideology of European superiority over that of other races primarily Blacks. One such case was that of Nott and Gliddon, which argued that the more elongated the human skull, the less inferior the individual was when compared to Europeans and apes. These elongated human skulls were often associated with Blacks. Gould further considered the case of Broca who fervently advocated that individuals who were of darker skin with a more elongated cranial structure were socially and intellectually inferior compared to individuals who had paler skin and had a

more proportioned cranial structure. The measurement of Blacks had less to do with their intellectual abilities and more to do with their physical traits. These examples, like many others in Gould's book, are egregious and continue to miss the mark on efficiently and effectively measuring the intellectual potential of Blacks compared to Whites. This includes concerns about Terrence's scores.

The value or lack thereof of measuring the intelligence humans is not only debated in research, but also in the media including televised programs. For example, in the 1970s classic *Good Times* episode, one learns about the story of Michael Evans who was being evaluated to determine his high school academic track (Reese, 2013). Although Michael was traditionally an A student, the results of IQ testing suggested that he would be unable to handle the academic rigor for those pursuing careers beyond trade school. Michael's family disappointingly received a letter recommending that he take a track more 'suitable' for students with lesser academic potential. Like this *Good Times* episode, many scholars and individuals in the field tend to rely heavily on test numbers, while failing to take into account the gestalt of the child including his or her physical environments, cultural norms and expectations, and lived experiences. Ford and Helms (2012) further determined that,

although there have been some remediation in testing disparities between Blacks and Whites, the unfair practices of assessments and testing negatively impact the hopes and dreams of Blacks.

Minority testing bias continues to persist in society and is often skewed to favor the educational and/or career advancement of Whites relative to Blacks (Berry, Clark, & McClure, 2011; Kwate, 2001). Unfortunately, the results of unfair testing and assessment practices are often used to make educational decisions that may or may not be in the best interest of the child. Stories like Terrence and that of Michael Evans continue to show that tests are not created fairly and equally for Blacks. Yet, there are some scholars who believe that the administration of IQ tests adequately and efficiently measures how successful an individual will be academically and beyond, as noted in Appendix A.

Terman, the reviser of the Stanford-Binet IQ test, is one such scholar who created and continues to create great tension with regard to intelligence testing and its utility in determining an individual's success rate. According to Terman (1916), IQ tests were deemed to accurately determine students' educational placements and supported intelligence tests as a basis for grading. Not only in the case of the retarded or exceptionally bright children, but with many

others also, intelligence tests can aid in correctly placing the child in school" (p. 16). He also believed that it was highly desirable that advancements or promotions be solely based on individuals' performances on IQ tests. Such erroneous biases pitting races against each other was simply a means to promote the dehumanization of Blacks and other non-European groups over the years using research that is still questionable (Gould, 1996). This practice continues.

The ongoing criticism remains that IQ tests are biased against culturally different groups, especially Blacks. In 1972, Williams challenged the authenticity and cultural fairness of traditional IQ tests when assessing Blacks. The birth of the Black Intelligence Test of Cultural Homogeneity (B.I.T.C.H) was designed as Williams believed that when Blacks were administered questions that were culturally relevant using familiar dialect, it would provide better input into how Blacks processed information. Using a sample size of 100 Black and 100 White high school students, the results of the administration of the B.I.T.C.H. revealed that Blacks outscored Whites at far more superior rate (Williams, 1972).

It is no wonder that scholars continue to question the validity and cultural fairness of traditional IQ tests among Blacks. Kwate (2001) asserted that traditional IQ tests are highly European centered and negate the equal representation of other minority groups within their sample sizes. Specifically, he stated that such tests, like the Wechsler Intelligence Scale for Children Third Edition (WISC-III) is "... a clear representation of the European rhetoric, in which Europeans are seen as universal" (p. 226).

Reynolds and Suzuki (2012) noted that, historically, on average, Blacks have always performed 1.0 standard deviation below Whites on most traditional IQ tests. Yet its critics continue to question if test bias among Blacks is real or simply a myth and excuse. Although there are some scholars who believe that test bias indeed exists in the performances between Blacks and Whites, many have explained it away as (1) being less of the actual standardized tests in and of themselves and (2) a function of those individuals administering such tests, regardless of race and ethnicity (Reynolds & Suzuki 2012; Sattler & Gywnne 1982). Instead, the issue lies in the individuals who are making the actual referrals for testing (Reynolds & Suzuki, 2012). In addition, with a 0.33 standard deviation reduction recently noted between Blacks and Whites performances on IQ tests, the IQ gap has decreased (Nisbett et al., 2012).

Contemporary Problems with
IQ Testing

Although there is a bleak history concerning the misuse and abuse of culturally biased IQ tests within minority and low socioeconomic communities, research consistently has shown that performances on IQ tests are not due to one's genetic makeup (Ang et al., 2010; Douglas, 2009; Eppig, 2010; Flynn, 2009; Trahan et al., 2014), but more closely aligned with an individual's acculturation to their environment and that of the mainstream population. Culture in itself implies that individuals will think, act, and feel in similar ways, thus different cultures have different beliefs on what constitutes intelligence (Ford, 2013; Sternberg, 2009). Tests that purport to measure the construct of intelligence will be culturally loaded based on the mainstream culture. Many test companies have tried attempts to correct for these biases by placing more culturally diverse relevant items on tests, increasing the number of individuals for culturally diverse backgrounds, and creating more nonverbal test items and tests (Sattler, 2001), but tests cannot be entirely eliminated of both linguistic and cultural influences (Flanagan, Ortiz, & Alfonso, 2013).

Despite this knowledge, the fields of psychology and education find themselves in the same predicament that they have always been in, which is educators making inappropriate and disproportionate referrals of minority students to special education and psychologists making harmful decisions based on faulty interpretations of the assessment data (Flanagan et al., 2013). Flanagan and colleagues further explained that although psychologists readily encounter individuals who have had differences in their opportunities to learn cultural knowledge or who lack English proficiency, this is often unaccounted for and not discussed within reports. Instead, vague statements are made, such as "results should be interpreted with caution" without any further investigation into whether the assessment was ever valid for those children to begin with, resulting in the same faulty and harmful decisions based on the inaccurate interpretation of the Full Scale IQ (Carpenter, Just, & Shell, 1990; Dorling, 2010; Flanagan et al., 2013; Ford, 2013; Naglieri & Ford, 2003).

The overall full scale IQ/composite is often made up of both the verbal and nonverbal abilities score. The problem with this overall score is that when these two scores are averaged together and are used to predict culturally different students potential for academic success, without taking into account discrepancies between the verbal and nonverbal scores or other cognitive abilities (Elliott, 1997; Thorndike, 1997), it

can result in these children being restricted from gifted programming and instead placed in classrooms with teachers who hold lowered expectations for them (Cartledge & Dukes, 2008; Ford, 2013; Jackson, 1999; National Association of School Psychologists, 2013; Nettles & Pleck, 1993; Oakes, 1995; Sullivan et al., 2009).

Many researchers and practitioners have advocated for the use of nonverbal measures of cognitive assessment with culturally and linguistically diverse children (Flanagan et. al., 2013; Ford, 2013; Naglieri & Ford, 2003; UNIT by McCallum & Bracken, 1997). The premise behind the use of nonverbal testing is that, by removing language, test scores are less influenced by culture, thus increasing cultural fairness in testing (Cummins, 2008; Noble, Tottenham, & Casey, 2005). In fact, some supporters of nonverbal intelligence tests for minority children have indicated that the use of such measures not only helps to prevent racially, culturally, and linguistically different children from being over-identified for special education programming, but also helps to ensure that these children have equal access to gifted education services (Ford, 2013; Naglieri, Rojahn, Matto, & Aquilino, 2005; Wiseley, 2001). While it is well acknowledged that nonverbal tests do reduce the level of cultural content on the test, others

in the field have also noted that the narrow range of items on many nonverbal tests may limit its utility in helping to effectively evaluate academic strengths and deficits (Flanagan et al., 2013). The example given by Flanagan and colleagues (2013) was that if a student has a reading weakness, the cognitive abilities of auditory processing (Ga)and crystallized knowledge (Gc) would be more useful in identifying deficits most related to the reading weakness, which are not assessed on tests of nonverbal abilities.

Best Practices

According to the American Psychological Association, IQ and achievement tests continue to be refined and improved in order to level the playing field for all racial and cultural groups. Several scholars including Howard Gardner and Robert Sternberg have suggested that IQ measures by themselves are insufficient in identifying the multifaceted nature of how individuals learn as well as their overall job success in the future (see APA, 2004). Therefore, a more holistic approach, which includes input from multiple sources and not solely IQ measures and achievement tests, are warranted in order to efficiently and effectively assess the gestalt of an individual.

The practice of basing student eligibility on a single discrepancy score contrasting intellectual ability and academic achievement continues to be used in schools around the United States. Due to the variability of minority students' cultural, linguistic, educational, and economic experiences, and the potential for cultural loading and test bias, the discrepancy model may not be the most appropriate method for use with students from diverse backgrounds. Rather, practitioners have begun to recognize the importance of analysis of multiple data sources to identify a student's academic areas of struggle and inform appropriate individualized interventions. Psychologists may find the cross-battery approach, the use of multiple assessment batteries coupled with a pattern of strengths and weaknesses, helpful in forming more comprehensive and reliable interpretations concerning a student's abilities than a singular-battery administration (Flanagan, Ortiz, & Alfonso, 2013; Schultz, Simpson, & Lynch, 2012).

It is especially important to assess students' strength across the seven higher-order broad CHC areas of cognitive functioning to include crystallized intelligence, fluid reasoning, visual-spatial ability, working memory, long-term retrieval, auditory reasoning, and processing speed (McGrew, 2005). Examining all broad areas

enables the practitioner to target a student's isolated areas of difficulty, and differentiate between normative and relative personal strengths and weaknesses. For example, cultural differences in learning may cause a student of diverse backgrounds to have greater difficulty on tests that heavily weigh social and language components and obtain scores indicative of a deficit in the areas of vocabulary and/or language, but his or her overall cognitive profile spans the verbal components, may fall well within the average range (Rice & Ortiz, 1994). Students whose cognitive profile reflects the aforementioned example may benefit from intensive interventions in areas of vocabulary, comprehension, and sequence and storytelling without identification as having a specific learning disability.

In addition, shifting from the use of a single global IQ score to a hierarchical model of intelligence may increase the percentage of racially, culturally and linguistically different students being identified for gifted special education programming (Pfeiffer, 2015). As stated above, the global IQ score is often an average of the verbal (crystallized intelligence) and nonverbal (fluid intelligence) score (Elliott, 1997; Thorndike, 1997). The problem with this is that even if the student does present with above average or superior abilities in one or more of the areas of

cognitive functioning, such as in the area of Fluid Intelligence, if he or she performs lower on one or more other areas, such as on tests of Crystallized Intelligence, he or she will most likely not meet the set criteria for the global ability cut off score. Cases like these occur often, especially for students who are economically, racially, culturally, and/or linguistically different from the mainstream. Based on these realities, one has to ponder why the global ability score is still so heavily relied on for placement decisions or, more so, why the global ability score is still relevant (Flanagan et al., 2013). The CHC allows us to move away from the global ability score and, instead, consider multiple areas of intelligence when making gifted identification decisions (2013, 2015).

Conclusion

Students who are racially, culturally, economically, and/or linguistically different from the mainstream culture (e.g., Terrence) continue to be misunderstood by their educators (teachers, administrators, counselors), resulting in numerous unnecessary referrals for special education services and limited referrals to gifted programming. Once these students are referred for special education or gifted evaluations, they are administered traditional psychological assessments that

are too often inappropriately interpreted to predict these students' learning potential. This practice of interpreting the Full Scale IQ composite as the best predictor of ability continues, despite the vast amount of research documenting that the definition of IQ is culturally derived based on the mainstream's culture and is not a fixed trait.

Recommendations for K-12 Teachers

1. In order to effectively differentiate the instruction when working with racially and culturally different students, teachers will need to adjust the curriculum, the classroom environment, content and activities, and overall work output in order to show students' mastery of the curriculum. This will ensure that the needs of all students are being met, including students like Terrence.

2. Teachers should integrate research-based academic instruction and classroom management techniques in the classroom that are culturally relevant and meets students' needs.

3. Teachers require support from district leaders in order to work with minority students. Ongoing professional development for teachers on culturally responsive pedagogy and student learning must include teaching strategies that involve developing compassionate

4. and respectful relationships with families and students of diverse cultures and backgrounds. Through embracing and celebrating differences, student and family overall responsiveness to learning and school engagement will increase.

5. Teachers should be sensitive and responsive to diversity and difference grounded in culture. This requires examination and acknowledgement of personal and professional biases when working with culturally diverse students, to ensure that these biases will not have negative implications for children and their families.

Recommendations for School Psychologists

1. School psychologists should choose assessments and test batteries based on the referral questions and the cultural appropriateness of the test batteries and not on the convenience of what test kit is readily available or easier to administer. When working with ethnic and language minority students, school psychologists should choose culturally sensitive tests that are less culturally/verbally loaded to assist with decision making for both gifted and special education programming.

2. School psychologists should aim to integrate data from the child's background history, such as level of acculturation, language acquisition, and education, when assessing students. School Psychologists should also integrate cognitive, achievement, and neuropsychological batteries using a cross battery assessment approach to more comprehensively address referral questions.

3. School psychologists should be careful not to further interpret the Full Scale IQ/composite score when large significant discrepancies exist across scales (e.g., verbal vs. nonverbal scales), because the full scale IQ is not meaningful and more likely to be misinterpreted. Instead, school psychologists should aim to fully communicate the student's strengths and weaknesses, carefully explaining how verbal ability is related to one's level of acculturation, and provide strength based recommendations to assist students with remediating their academic weaknesses.

Recommendations for Families

1. Parents should be aware that the results of an IQ score provides a numerical score that offers insight as to how students perform compared to same age and grade peers. IQ scores do not predict future academic and/or job success. Therefore, parents should continue to educate themselves on other ways students are smart (e.g., Howard Gardner's Multiple Intelligence).

2. Parents should ensure that the results of assessments are explained to them in a manner that is free from technical jargon. Specifically, there should be a clear indication as to how the results of assessments impact the students' learning both at school and at home. This will assist parents in finding other strategies to assist with their child(ren) at home as well as select appropriate outside learning strategists to remediate student learning when needed.

3. Parents should realize that all children learn at a different rate and pace. IQ is very fluid and may present high or low scores during the younger years than the latter years. Therefore, parents should focus on the patterns of strengths and weaknesses identified in the assessment rather than the numerical score as this information is critical in understanding how their child learns best.

References

Ang, S., Rodgers, J. L., & Wanstrom, L. (2010). The Flynn effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-children data. *Intelligence, 38*(4), 367-384.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, *96*(5), 881-906.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since "bias in mental testing." *School Psychology Quarterly*, *14*(3), 208–238.

Cartledge, G., & Dukes, C. (2008). African American children in special education: Definition and dimensions. *Current Issues, 24*(5), 383-395.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404-431.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In N. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 487-499). New York: Springer.

Dorling, D. (2010). The return to elitism in education. *Soundings*, *44*(1), 35-46.

Douglas, B. (2014, January 3). Race, intelligence, and IQ. Are Blacks smarter than Whites? Retrieved October 28, 2016, from the Africa Resource website: http://www.africaresource.com/

Elliott, C. D. (1997). The differential abilities scales. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 517-532). New York: Guilford Press.

Eppig, C. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3801-3808.

FairTest fact sheets [Fact sheet]. (n.d.). Retrieved October 28, 2016, from the FairTest website: http://www.fairtest.org/

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken: John Wiley.

Flynn, J. R. (2009). The WAIS-III and WAIS-IV: Daubert motions favor the certainly false over the approximately true. *Applied Neuropsychology, 16*(2), 98-104.

Ford., D.Y. (2013). *Recruiting and retaining culturally different students in gifted education*. Waco: Prufrock Press.

Ford, D. Y., & Helms, J. E. (2012). Overview and introduction: Testing and assessing African Americans: "Unbiased" tests are still unfair. *Journal of Negro Education*, *81*(3), 186-189.

Ford, D. Y., & Russo, C. J. (2016). Historical and legal overview of special education overrepresentation: Access and equity denied. *Multiple Voices for Ethnically Diverse Exceptional Learners*, *16*(1), 50-57.

Gould, S. J. (1996). *The mismeasure of man*. New York: WW Norton.

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Intelligence and achievement testing: Is the half full glass getting fuller? (2004, June 17). Retrieved October 28, 2016, from the American Psychological Association website: http://www.apa.org/

Jackson, J. F. (1999). What are the real risk factors for African American children? *Phi Delta Kappan, 81*(4), 308-312.

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). *The Effects of Standardized Testing*. Dordrecht, Netherlands: Springer.

Kendi, I.X. (2016, October 20). Why the achievement gap is a racist idea. Retrieved October 28, 2016, from the AAIHS website: http://www.aaihs.org/

Kwate, N. O. A. (2001). Intelligence or misorientation? Eurocentrism in the WISC-III. *Journal of Black Psychology, 27*(2), 221-238.

McCallum, R.S. & Bracken, B.A. (1997). The universal nonverbal intelligence test. In D.P. Flanagan, J.L. Genshaft, & P. L. Harrison (Eds.), *Beyond traditional intellectual assessment: Contemporary and emerging theories, tests, and issues*. New York: Guilford Press.

McGrew, K. S., (2005). The Cattell-Horn-Carroll Theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. (2nd ed., pp. 136-181). New York: Guilford Press.

Mercer, C. D., & Pullen, P. C. (2009). *Students with learning disabilities*. Upper Saddle River, NJ: 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*(2), 155-160.

Naglieri, J. A., Rojahn, J., Matto, H. C., & Aquilino, S. A. (2005). Black-white differences in cognitive processing: A study of the planning, attention, simultaneous, and successive theory of intelligence. *Journal of Psychoeducational Assessment, 23*(2), 146-160.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130-159.

Noble, K. G., Tottenham, N., & Casey, B. J. (2005). Neuroscience perspectives on disparities in school readiness and cognitive achievement. *The Future of Children*, *15*(1), 71-89.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Ortiz, S. O., & Ochoa, S. H. (2005). Advances in cognitive assessment of culturally and linguistically diverse individuals. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (234-250). New York: Guilford Press.

Pfeiffer, S. I. (2015). Tripartite model of giftedness and best practices in gifted assessment. *Revista de Educación*, *368*, 155-182.

Racial and ethnic disproportionality in education [Position statement]. (2013). Retrieved October 28, 2016, from the National Association of School Psychologists website: http://www.nasponline.org/

Reese, R. (2013, February 21). Minority testing bias persists. Retrieved October 28, 2016, from the Huffington Post website: http://www.huffingtonpost.com/

Reynolds, C. R. (1982a). Construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 199–227). Baltimore: Johns Hopkins University Press.

Reynolds, C. R. (1982b). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Wiley.

Reynolds, C. R., & Brown, R. T. (1984). *Perspectives on bias in mental testing*. New York: Plenum Press.

Reynolds, C. R., & Suzuki, L. A. (2013). Bias in psychological assessment: An empirical review and recommendations. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.), *Handbook of Psychology: Assessment psychology* (2nd ed., Vol. 10, pp. 82-113). Hoboken: Wiley.

Rice, L. S., & Ortiz, A. A. (1994). Second language learning difference or learning disability? *LD Forum, 19*(2), 11-13.

Salvia, J. & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9th ed.). Boston: Houghton Mifflin.

Sattler, J. M. (2001). *Assessment of children: Cognitive foundations* (4th ed.) San Diego: Jerome M. Sattler.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.) San Diego: Jerome M. Sattler.

Sattler, J. M., & Gwynne, J. (1982). White examiners generally do not impede the intelligence test performance of Black children: To debunk a myth. *Journal of Consulting and Clinical Psychology, 50*(2), 196-208.

Schultz, E. K., Simpson, C. G., & Lynch, S. (2012). Specific learning disability identification: What constitutes a pattern of strengths and weaknesses. *Learning Disabilities, 18*(2), 87-97.

Sullivan, A. L., A'Vant, E., Baker, J., Chandler, D., Graves, S., McKinney, E., & Sayles, T. (2009). Inequity in special education, part I: Understanding the problem of disproportionality. *Communiqué, 38*(1), 14–15.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale.* Boston: Houghton Mifflin.

Thorndike, R. (1997). The early history of intelligence testing. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-17). New York: Guilford.

Trahan, L. H., Stueling, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: *A meta-analysis. Psychological Bulletin, 140*(5), 1332-13360.

Williams, R. L. (1972, September). *The BITCH-100: A culture-specific test.* Paper presented at the American Psychological Association Annual Convention, Honolulu. Retrieved from ERIC database. (ED070799)

Wiseley, M. C. (2001). *Non-verbal intelligence and Native-American Navajo children: A comparison between the CTONI and the WISC-III* (Unpublished doctoral dissertation). University of Arizona, Tucson.

Donna Y. Ford, Ph.D., is a professor in the Department of Special Education and the Department of Teaching and Learning at Vanderbilt University. All of her degrees are from Cleveland State University (B.A. in communications and Spanish; M.Ed. in counseling, and Ph.D. in Educational Psychology with a focus on urban education and gifted education). She conducts research and writes extensively on: (1) the achievement gap; (2) recruiting and retaining culturally different students in

gifted education; (3) multicultural curriculum and instruction; (4) culturally competent teacher training and development; (5) African-American identity; and (6) African-American family involvement. Professor Ford consults with school districts, and educational and legal organizations on these areas. Dr. Ford has authored some 200 articles and chapters, and more than eight books. She can be contacted at donna.ford@vanderbilt.edu

April J. Lisbon, Ed.D., is a school psychologist currently employed with the District of Columbia Public Schools. Her current research interest includes special education advocacy as related to the over identification of minority groups in special education and twice exceptional students. She can be contacted at dr.lisbonpeoples@yahoo.com.

Natasha N. Little-Harrison, Psy.D., is a licensed clinical and school psychologist currently in private practice in Richmond, Virginia Her current research interests include promoting student achievement through the realms of positive psychology. She can be contacted at Dr.nlharrison@richmondchildpsychologist. com.

Appendix A presents a comparison of supporters and opponents of testing issues, which implications for Black students.

Issue/Topic	What test supporters say ...	What test critics say ...
Fairness, Civil rights	Testing all students is the best way to measure how effective schools are, and that state or local content standards ensure that all students are learning the same curriculum. Students disadvantaged by tests can be better served by holding their schools accountable when they perform poorly on tests.	Tests can contain culturally biased content that may be unfamiliar to minorities and recent immigrants. Moreover, for students with learning disabilities or who process information differently, the nature of the test itself (be it multiple choice or short answer format) may be unfair. Tests do not adequately measure student and school performance, and that judging (and in some cases punishing) schools with low test scores results in even fewer resources for the students who need them most.
What the tests measure	Developing and administering tests that measure students' knowledge against learning standards will ensure that all students have certain proficiencies and are not left behind or falsely promoted from grade to grade.	Many tests created for national use may not include content emphasized at the state level, resulting in students being tested on material they have not been taught. A "narrowing" of the curriculum, saying that the heightened attention paid to standardized tests forces teachers to ignore content or even entire subjects that do not appear on the tests.

Issue/Topic	What test supporters say …	What test critics say …
The use of "high-stakes" tests	Tests are an important part of "raising the bar" on student performance. Attaching test results to grade promotion, graduation, and teacher evaluation will send a strong message to students, teachers, school leaders, and parents that students must meet proficiency levels.	"Test anxiety" and stereotype threat may affect a student's performance, resulting in scores that do not adequately reflect his or her knowledge. Standardized tests are only one measure of student performance, and must be considered alongside other assessment tools, including classroom work, student portfolios, and teacher evaluations.
The validity of test scores	Standardized tests are the most objective and accurate assessments of students' knowledge and skills. By creating norm groups or specific criteria to which students are compared, test makers can measure each student's abilities with precision.	Test-making is far from a perfect science. Tests may contain errors, making results inconclusive. Furthermore, a test's standard error of measurement may be large enough to call into question the accuracy and usefulness of the results.
Using tests to determine school funding	Schools should be rewarded financially for performing well on standardized tests, and that providing such incentives will motivate school leaders and teachers to teach effectively and raise student performance.	Financial rewards for schools in which students perform well is an inappropriate use of funds. It is unfair to expect students at schools in impoverished areas to perform as well as those in wealthy areas, and withholding additional funding for schools in need will stagnate performance levels.