Misconceptions About the Naglieri Nonverbal Ability Test: A Commentary of Concerns and Disagreements

Jack A. Naglieri & Donna Y. Ford

To cite this article: Jack A. Naglieri & Donna Y. Ford (2015) Misconceptions About the Naglieri Nonverbal Ability Test: A Commentary of Concerns and Disagreements, Roeper Review, 37:4, 234-240, DOI: 10.1080/02783193.2015.1077497

To link to this article: http://dx.doi.org/10.1080/02783193.2015.1077497
Misconceptions About the Naglieri Nonverbal Ability Test: A Commentary of Concerns and Disagreements

Jack A. Naglieri and Donna Y. Ford

Black and Hispanic students are undeniably underidentified as gifted and underrepresented in gifted education. The underrepresentation of the two largest groups of “minority” students is long-standing, dating several decades, and is a serious area of contention. Most debates focus on the efficacy of traditional intelligence tests with verbal, quantitative, and nonverbal scales compared to intelligence tests that are nonverbal when identifying underserved gifted students. Test developers; researchers; federal, state, and local government officials; policymakers; administrators; and educators have debated different solutions to the problem of underrepresentation of minorities in gifted educational programs for decades. Controversies surrounding how to equitably identify these gifted students abound, and arguments are quite polemic and entrenched; nonetheless, in many instances gifted Hispanic and Black students are often disproportionately denied access to gifted education because of the methods and instruments used. In this article, we review a study by Giessman, Gambrell, and Stebbins regarding one Nonverbal Test of General Ability (NNAT2), which has been widely used for identification of gifted non-White students. We address concerns about conclusions raised by Giessman and coauthors and present cautions about the problems involved in reporting archival data.

Keywords: Black students, CogAT, gifted identification, gifted underrepresentation, Hispanic students, NNAT, nonverbal tests, test fairness

In a recent research-based article, Giessman, Gambrell, and Stebbins (2013) examined several aspects of the Naglieri Nonverbal Ability Test–Second Edition (Naglieri, 2008a) in order to evaluate the scores it yields across race and ethnicity. Giessman et al. (2013) compared the Naglieri Nonverbal Ability Test (NNAT2; Naglieri, 2008a) to the Cognitive Ability Test (CogAT6; Lohman & Hagen, 2001) for identifying Black and Hispanic students. Giessman et al.’s (2013) main claim is that the NNAT2 does not address the issue of underrepresentation in gifted programs as noted, for instance, by Naglieri and Ford (2003) and Naglieri, Brulles, and Lansdowne (2009). In this response, we point out and consider flaws in their methodology and the need for more information about the CogAT6 to better understand how best to give Black and Hispanic students equitable access to gifted education.

Ford (e.g., 2010, 2013) has written extensively on the factors related to underrepresentation of Black and Hispanic students in gifted education. She and Naglieri (Naglieri & Ford, 2005) have reported that federal, state, and district data from the Office for Civil Rights clearly document that Black and Hispanic students are not only underrepresented in gifted education but also that their poor presence is inequitable. Using the Office for Civil Rights Civil Rights Data Collection, with 2011–2012 as the most recent case in point, Black students represent 19% of U.S. public school students but only 10% of students in gifted education; this is almost a 50% discrepancy. Hispanic students comprise 25% of public school students but only 16% of gifted education, representing about a 40% discrepancy. Every year that data have been collected, the Civil Rights Data Collection reports (various years between 2004 and 2012) indicate that Black students are the most underrepresented racial group, followed by Hispanic students. Keeping inequity and underrepresentation in mind, some professionals in the gifted education field have emphasized the need for reform of the methods (policy and procedures) and measures (tests and
checklists or forms) used to identify all gifted students (see Ford, 2013; Naglieri, 2008b).

We agree with the recognition by Giessman et al. (2013) that the problem of Hispanic and Black underrepresentation in gifted education must be addressed. This important longstanding identification, access, and service problem is especially prevalent with Black and Hispanic students who have always been significantly unidentified as intelligent and highly capable in our gifted education system. Ford (2013) detailed a litany of inequitable barriers, ranking traditional ability tests with verbal and quantitative content as a primary and impactful obstacle. Accordingly, mainly Black and Hispanic students have been and remain extensively underrepresented in gifted education (Baldwin, 2004; Castellano & Frasier, 2010; Ford, 2013; Ford, Grantham, & Whiting, 2008; Frasier, Garcia, & Passow, 1995; Office for Civil Rights, 2004, 2006, 2009, 2011, 2012). This access and equity concern is consistent with the National Association of Gifted Children’s (2010) recommendation regarding the need to better address testing, identification, and underrepresentation barriers, as well as that of professional organizations (e.g., Council for Exceptional Children, American Psychological Association, and American Educational Association). Access to gifted education is also at the core of our individual and collective works on the equitable testing, assessment, and identification of Hispanic and Black students using nonverbal tests.

Our position is clear—the need to identify and serve more non-White students for gifted programming is not just an important issue for educators, it is a critical social justice issue that has considerable implications for our country. To state the obvious, when an increasingly large percentage of our school population (namely, Black and Hispanic students) goes unidentified and underidentified as gifted and is in need of intellectual and academic challenge, our schools lose nationally and internationally, as we learned several decades ago with Sputnik and continue to learn with international research reports. Our nation cannot afford to underidentify and misidentify minority students who are almost the majority of our K–12 students (e.g., U.S. Department of Education, 2014). What is the field of education, and gifted education specifically, resisting and/or waiting for in terms of being equitable, proactive, and culturally responsive to Black and Hispanic students (Darling-Hammond, 2010)?

**DEFINING GIFTED**

Achieving the goal of equitable representation of Black and Hispanic students in gifted educational programs is based not only on the ways in which gifted students are identified but also the very definition of *gifted*, which should consider culture, cultural characteristics, and cultural differences (Castellano & Frasier, 2010; Ford, 2010, 2013; Frasier et al., 1995; Sternberg, 2007a, 2007b). Importantly, the definition of gifted has a profound influence on the tools used to identify these non-White students, particularly when the goal is to decrease and preferentially eliminate gifted underrepresentation among these students. For example, Naglieri (2008b) defined gifted students as those with very high “general ability” measured with a nonverbal (or other ability test), and Giessman et al. (2013) chose “academic giftedness” (p. 108). This means that students who are high in ability and potential but may not be currently achieving at high levels in school would not likely receive the services needed to bring their achievement up to the level of their general ability. This is a major and meaningful difference in perspectives that has important and far-reaching implications for professionals who are responsible for identifying gifted students (see Naglieri, 2008b, for more discussion) and ensuring access for underrepresented groups. Requiring that a student be “academically gifted” means that academic skills are already present or being demonstrated; however, requiring high general ability means that the student has the capability to be high achieving, given support and opportunity to learn.

The purpose of a nonverbal measure of general ability is to measure “ability” with tests that do not require verbal, social, and quantitative knowledge, especially considering that the aforementioned minority students often have not had the opportunity to acquire the academic information these tests require to the same degree as White students, especially those who are higher income, academically privileged, and socially advantaged (Ford, 1994; 2013; Helms, 1992; Naglieri, 2008b). At the root of this approach is the expectation that there would be a smaller difference between racial and cultural groups on a nonverbal measure of general ability than on a traditional verbal and quantitative measure of general ability.

Support for the previous claim that a nonverbal measure, like the NNAT/NNAT2 (Naglieri, 1997, 2008a), provides scores that are similar across groups was initially reported by Naglieri and Ronning (2000). They compared racial and ethnic groups by creating three samples (White/African American; White/Hispanic; White/Asian) matched by geographic region, socioeconomic status, ethnicity, and public/private school (N = 22,620). They found small differences in NNAT scores for each of the three comparisons. This was followed by Naglieri and Ford’s (2003) study, which reported similar rates of identification across groups. The next study, by Naglieri, Booth, and Winsler (2004), compared NNAT scores for Hispanic children with limited English proficiency (n = 148) and those with adequate English skills (n = 148) using groups selected from the NNAT standardization sample (N = 22,620) and matched on geographic region, gender, socioeconomic status, urbanicity, and race/ethnicity. They found a very small difference (d ratio = 0.1) between the NNAT standard scores for the children with limited English proficiency (M = 98.0) and those without limited English proficiency (M = 96.7); however, significant differences were found between these groups on measures of vocabulary, reading comprehension, and
listening comprehension. The implication of that study was that language-based tests pose an obstacle to gifted placement for students with limited English-language skills.

Naglieri and Ford’s (2003) paper was criticized by Lohman (2005; author of the CogAT, which is a competing measure) on the basis of psychometric issues. Naglieri and Ford (2005) responded to Lohman by stressing that a nonverbal measure of general ability is a critical and equitable component of the recruitment (identification and assessment) process for gifted Black and Hispanic students who are intellectually or academically gifted. In a 2013 court case, discussed later, the NNAT was found to be effective with Hispanic students (see McFadden v. Board of Education, 2013).

Specifically, it is essential to distinguish between students with high general ability on a nonverbal test, regardless of their verbal or quantitative skills, versus students who may be academically gifted (i.e., high scores on verbal and/or quantitative tests). Some (e.g., Carman & Taylor, 2010; Giessman et al., 2013) have echoed Lohman’s concerns when standard deviations for local samples did not match the values used in norming. This will be more fully discussed later, but first is the need for clarification about the use of verbal, quantitative, and nonverbal tests and/or subscales.

A test of intellectual ability that requires vocabulary and quantitative skills (e.g., CogAT6) could be an appropriate measure of general ability for those students with equal opportunities to acquire knowledge of words and math. These tests, however, are inappropriate measures of general ability for those students/test takers who have limited knowledge of the language or math required by the test questions, which is a function of exposure to social and educational experiences (e.g., Prifitera, Saklofske, & Weiss, 2005). If a test of intellectual ability requires knowledge of words, and a child (e.g., due to race, ethnicity, language, income, school quality) has had limited opportunity to learn words, then the score the ability test yields will not be an accurate reflection of ability or potential. Scores on verbal and quantitative tests for someone with limited educational, social, and economic opportunities should not be interpreted as ability but rather as language and math skills at the time of testing. This concept was originally proposed by Yoakum and Yerkes (1920), who stated that the verbal and quantitative tests were intended for “men who can read and write English fairly well (illiterates)…” and the nonverbal measures were designed to be used for “men who are unable to read and write English well (illiterates)” (p. 15). The verbal/quantitative and nonverbal distinction was designed to “minimize the handicap of men who because of foreign birth or lack of education are little skilled in the use of English” (p. 17) and “in order that injustice by reason of relative unfamiliarity with English may be avoided” (p 19). The equitable and culturally responsive purpose of a nonverbal measure of general ability, therefore, has been to measure general ability without the confounding influence of knowledge, access, opportunity, language, income, cultural differences, and other inequities confronting Black and Hispanic students (Kaufman, 1994, 2009), including educators’ low expectations and deficit thinking (e.g., Ford, 2010). Importantly, these tests were not described as measures of verbal, quantitative, or nonverbal abilities.

Group ability tests like the CogAT6, as well as individual measures such as the Wechsler Intelligence Scale for Children (WISC; Wechsler, 2003), measure general ability using questions that require (a) verbal knowledge, (b) quantitative knowledge, and (c) solution of problems presented using diagrams and figures that, although they can have verbal labels (e.g., square, circle, etc.), do not require verbal knowledge. The interpretation of these various tests was clarified by Kaufman (2006):

Although his [David Wechsler’s] intelligence tests in the 1930s and 1940s departed from the one-score Stanford-Binet by offering separate Verbal and Performance [nonverbal] IQs as well as a profile of scaled scores, Dr. Wechsler remained a firm believer in Spearman’s ‘g’ theory throughout his lifetime. He believed that his Verbal and Performance Scales represented different ways to access ‘g’, but he never believed in nonverbal [or verbal] intelligence as being separate from ‘g’. . . . He saw the Performance Scale as the most sensible way to measure the general intelligence of people with hearing impairments, language disorders, or limited proficiency in English. (p. iv)

It is important to recognize, however, that including a nonverbal measure of ability does not guarantee a more appropriate and equitable representation of Black and Hispanic students in gifted education, but it can increase the probability of such students coming to the attention of decision makers. Take, for example, the case of McFadden vs. Board of Education (2013) where Hispanic and Black students were significantly underrepresented in gifted education classes. Although over 40% of the students in Elgin school district (U-46) were Hispanic, only 2% of the students in the district’s mainstream elementary (Grades 4–6) school gifted program were Hispanic. To enter the mainstream gifted program, the CoGAT was adopted; for the former English-language learner (ELL) students in the nonmainstream gifted program, the NNAT was adopted. Hispanic students represented all students in the nonmainstream gifted program.

The district separated students into different gifted classes based on former ELL status and required them, for admission to the mainstream gifted program, to obtain high scores on a measure of ability that demands knowledge of English and quantitative skills (e.g., the CogAT6; Lohman & Hagen, 2001). The Court ruled that the use of the CogAT6 contributed to the underrepresentation of Hispanic students in gifted education. In this contemporary and unprecedented case, the NNAT was deemed effective and equitable at increasing access to gifted education for Hispanic students in the elementary nonmainstream gifted program. At middle
school, the NNAT was no longer used for gifted education identification. Hispanic students, but not their White classmates, were retested. When the CogAT6 was used and the NNAT was dropped, Hispanic students were woefully underidentified and underrepresented in gifted middle school programs in U-46.

Judge Gettleman wrote that “there is no question that the District placed gifted Hispanic students [in different programs] based solely on their cultural identity.” According to Judge Gettleman, a separate, segregated program that discriminates against Hispanic students violates the Equal Protection Clause of the 14th Amendment to the United States Constitution and the Illinois Civil Rights Act. This case illustrates that though including a nonverbal test alone will not solve the problem of underrepresentation, it could, if used correctly, increase access for Hispanic students born in the United States, Hispanic students who are ELL, and Black students (Ford, 2013).

METHODOLOGICAL ISSUES RELATED TO TEST SCORE COMPARISONS

It is our belief that comparing test scores across groups, regardless of whether the comparison is based on age, grade, income, gender, race/ethnicity, or English-language skills, should be conducted with samples that are as similarly matched as possible. Matching is used to reduce the influences of variables other than the variable(s) of interest. In some cases, logistical constraints limit the extent to which similar groups can be found. Other times, samples can be trimmed or subjects selected expressly to get similar groups to provide for a better examination of differences in scores. For example, Naglieri et al. (2004) selected cases from the NNAT standardization sample (N = 22,620) that were matched on geographic region, gender, socioeconomic status, urbanicity, and ethnicity when they compared scores for Hispanic children with (n = 148) and without (n = 148) limited English proficiency. They found a very small difference (d ratio = 0.1) between the NNAT standard scores for the children with limited English proficiency (M = 98.0) and those without limited English proficiency (M = 96.7), but they also found significant differences between these groups on measures of Vocabulary, Reading Comprehension, and Listening Comprehension of the Stanford Achievement Test. The use of careful matching showed that English proficiency was related to group differences but ability, measured using a nonverbal measure of general ability, was not.

Note that Giessman et al. (2013), in contrast to other published research on race and ethnic differences (e.g., Naglieri, Rojahn, & Matto, 2007; Naglieri, Rojahn, Matto, & Aquilino, 2005), did not match the groups or statistically control for demographic differences. This calls into question the validity of their findings and interpretations because the results are confounded by the variance accounted for by different characteristics of the subjects in the samples. Giessman et al. (2013) compared results for different samples administered the CogAT6 and NNAT2, not recognizing that sample differences likely impacted test score differences. In fact, the sample used to obtain CogAT6 results had 1,798 more students than the one used to obtain NNAT2 results. Just as important, the samples were given these tests several years apart. A better methodology could have been utilized by only using cases where both tests were given to the same students. We are also concerned that the data were not obtained using a counterbalanced administration order of the CogAT6 and NNAT2 to control for practice effects, adding more threats to the validity of their study. These methodological problems likely had a considerable impact on the findings, especially the reported means and standard deviations.

METHODOLOGICAL ISSUES RELATED TO MEANS AND STANDARD DEVIATIONS

Giessman et al. (2013) examined the means and standard deviations obtained for different samples of students who were given the NNAT2 and CogAT6. They reported higher CogAT6 mean scores than the NNAT2 by about 6 points, yet no statistical testing of the difference was conducted. What they omit is an explanation for this difference, the most obvious being that, due to the Flynn effect (Flynn, 1987), this difference is likely the result of softening of norms for the CogAT6. That is, older norms have consistently yielded inflated scores, which is probably one of the reasons that the CogAT6 has been revised.

Giessman et al. (2013) stated that NAI score standard deviations were “larger than expected 16” (p. 105) for the NNAT2 (N = 4,035), but it is not reasonable to expect that the standard deviations on any test should be the same as the normative values because there is no evidence whatsoever provided by Giessman and coauthors that their samples match the U.S. population. In fact, it is clear that the sample used by Giessman et al. should not be used to compare standard deviations in the local samples to the one used for norming the NNAT2 because:

- the sample is restricted to one school district;
- no calibration of socioeconomic status was reported; and
- the distribution by ethnicity was considerably discrepant from the U.S. population values (e.g., the U.S. population of Hispanics is 20.1% [see Naglieri, 2012] but only 5% of their sample was Hispanic).
The fact that their sample does not resemble the U.S. population means that it is unreasonable to compare the obtained standard deviations of the NNAT2, as well as the CogAT6, to the values prescribed in their respective test manuals. In other words, it is not the case that studies of any standardized test using samples that are not matched to the U.S. population and are not sufficiently large will yield a mean and standard deviation equivalent to the values the test was calibrated to.

In addition, because the NNAT2 is normed by age and the test administered by level, it is unreasonable to expect standard deviations of 16, regardless of the size of the sample. There is a large range in age in any grade and that variability increases the standard deviations. Importantly, however, in the most recent study of the NNAT2 (Naglieri, 2011) it was reported that for the data used for the Normative Update, the mean score was 100 and standard deviation 16 for the entire normative sample (N = 99,004). Users of the NNAT2 (or any test) should expect that samples from one school district, no matter how large they may be, should not yield means and standard deviations that match the values set during norming.

METHODOLOGICAL ISSUES RELATED TO IDENTIFICATION RATES

The data used by Giessman et al. (2013) would have been more informative if their analysis of NNAT2 and CogAT6 included identification rates for all the scores on these two tests, including the Verbal and Quantitative portions of the CogAT6. The influence these scores have when used as part of the identification processes cannot be denied—very few Black and Hispanic students would earn scores high enough to be identified, and this was evident in McFadden vs. Board of Education (2013). Giessman et al.’s (2013) emphasis on the NNAT2 versus the CogAT6 Nonverbal subscale is important, but it should include an analysis or analyses of the separate contribution the verbal and quantitative scales make. If school administrators have the three CogAT6 subscale scores, all of which are described as different measures of ability, and these sub-scores are used in some kind of matrix system, then students with limited English-language skills and many students from traditionally underrepresented populations are not likely to be identified. Additionally, if the total CogAT6 score is used, that will also exclude capable students from underrepresented groups as evidenced by Giessman et al.’s (2013) finding that, at the 10% level (presumably this means the 90th percentile rank, although this is not explicitly stated in their manuscript), 10.4% of students with limited English-language skills were identified by the NNAT2 but only 1.0% were identified by the CogAT6 total score, which includes all three scales.

METHODOLOGICAL ISSUES RELATED TO THE WISC-IV

Giessman et al. (2013) tested the validity of the CogAT6 and NNAT2 by comparing the mean scores earned by students in the top 5% of the different samples of children. They argue that the combined score from the CogAT6 (Verbal, Quantitative, and Nonverbal) was a better predictor of the WISC-IV than the NNAT2. This comparison is problematic for several reasons but especially because the NNAT2 scores do not have the same content as the academically laden verbal and quantitative measures that are included in both the CogAT6 and the WISC-IV. That is, the point of the NNAT2 is to measure general ability without the impact of language and educational skills on the measure of general ability. What is apparent in the data they provided is that the NNAT2 and the WISC-IV PRI mean scores were very similar. This is an important finding because the “NAI appeared nominally better at predicting high PRI, but the difference was not significant” (Giessman et al., 2013, p. 106). Additionally, in order to examine the predictive validity of the NNAT2 and CogAT6 compared to any other ability test, using an unrestricted sample would have been considerably more effective and would have avoided the problem of restriction in range apparent in the small standard deviations (see Giessman et al., 2013, table 4). The data that they have could have been more informative had the entire sample, rather than only the top 5%, been examined.

CONCLUSIONS

Our goal in this response was to restate that decreasing and/or eliminating the underrepresentation of Black and Hispanic students in gifted education programs must be a fundamental goal of educators (teachers, administrators, school boards), test administrators (e.g., school psychologists), test developers, as well as policymakers and decision makers. Achieving this equitable and culturally responsive goal has been hampered by, for example, disagreements about the very definition of gifted and the tests used in the identification process. We strongly argue that the use of tools that require verbal and quantitative knowledge perpetuates conceptualizing students within an academic giftedness framework. To be clear, however, we have no objection to finding academically gifted students or using verbal and quantitative measures as part of the gifted identification process—as long as these measures do not block the participation of otherwise smart and potentially capable Hispanic and Black students in gifted education. We believe that accurate and equitable evaluation of underrepresented groups can be achieved by using nonverbal tests of general ability as a part of the process to identify gifted students who are capable, and with adequate instruction they will attain a high level of academic achievement. We urge our colleagues to...
recognize that our system of identification has failed to give Hispanic and Black students fair treatment, resources, and opportunities that an appropriate education could offer, and we need to overcome this injustice.

REFERENCES

AUTHOR BIOS

Jack A. Naglieri, PhD, is Research Professor at the Curry School of Education at the University of Virginia, Senior Research Scientist at the Devereux Center for Resilient Children, and Emeritus Professor of Psychology at George Mason University. Dr. Naglieri is the author or coauthor of numerous scholarly papers, books, and tests. His scholarly research includes investigations related to topics such as intellectual disabilities, specific learning disabilities, giftedness, and attention deficit disorder; psychometric studies of tests such as the Wechsler Scales of Intelligence, Naglieri Nonverbal Ability Test, Cognitive Assessment System, and the Kaufman Assessment Battery for Children; examination of race, gender, and ethnic differences in cognitive processing; fair assessment using nonverbal and cognitive processing tests; identification of gifted minorities; IDEA and identification of specific learning disabilities; and cognitively based academic interventions. E-mail: jnaglieri@gmail.com

Donna Y. Ford, PhD, is a professor of special education in the College of Education at Vanderbilt University. She is the author of several books, chapters, and articles that focus on underrepresentation among gifted Black and Hispanic students. Her scholarly foci include closing the achievement gaps; underachievement among gifted Black students; identification and assessment of gifted underrepresented students; addressing social, emotional and psychological development of culturally different students; multicultural counseling; and multicultural and culturally responsive education for gifted students. Dr. Ford consults nationally on recruiting and retaining underrepresented students in gifted education and supporting educators in becoming culturally competent. She is a member of several professional organizations. E-mail: donna.ford@vanderbilt.edu