

Group Differences on the Naglieri Nonverbal Ability Test (NNAT)

George Manos^{*}

Baker County School Board, Florida

Abstract

The claim of the author of the NNAT is that it produces equal proportions of Black, White, Hispanic, and low-income students scoring in the gifted range. The present study involved 392 second grade students from a rural Florida school district who were administered the NNAT in a group format. The testing by the school district was for the purpose of increasing the number of low-income and minority students in their Gifted Program. A cut-score of 115 was used as the criterion for convening a child study team meeting to discuss if a formal intellectual evaluation by the district school psychologist was warranted. Of the total sample, only the Black, White, and low-income subsamples had sufficient students to allow for statistically meaningful comparisons. Mean IQ scores for Blacks (89.0) and Whites (98.1) differed by 9.1 standard score points. Students receiving a free or reduced price lunch (used as a proxy measure for low socioeconomic status) were compared with students not receiving this service. Mean IQ scores for all students on free or reduced lunch status (93.5) and all students not receiving this service (100.6) differed by 7.1 standard score points. Mean IQs of White students receiving a free or reduced price lunch (95.1) and White students not receiving this service (100.6) differed by 5.5 standard score points. Matching Black and White students on free and reduced lunch status reduced the Black-White difference by approximately one-third, or a difference of 6.0 standard score points (89.1 and 95.1, respectively). The results of this study do not support Naglieri's claim.

Key Words: IQ; NNAT; Blacks; Whites; Minorities; Low income groups; Naglieri; Gifted; Disproportionality.

The use of nonverbal tests of intelligence for the identification of intellectually gifted low-income and minority students has been advocated by those who consider such tests to be less biased towards culturally, linguistically and racially diverse populations. A further impetus for the use of such tests is the under-representation of the aforementioned groups in public school programs for the intellectually gifted and the concomitant belief that the use of such tests will ameliorate the problem of "disproportionality." Tests such as the Kaufman Assessment Battery for Children (KABC) and the more recent Kaufman Assessment Battery for Children-Second Edition (KABC-II), the Universal Test of Nonverbal Intelligence (UNIT), the Comprehensive Test of Nonverbal Intelligence (CTONI), and the Naglieri Nonverbal Ability Test (NNAT) are examples of nonverbal tests that are touted as less biased towards minority and low-income groups than are the more verbally weighted intellectual measures. Such tests show reduced group differences, in contrast to traditional intelligence tests, such as the Wechsler series and the Stanford-Binet: Fourth Edition (SB:FE), which show an approximately one standard deviation difference between Blacks and Whites at the adult level (Gottfredson, 1997).

Proponents of nonverbal tests tend to define bias not in the conventional statistical sense, where bias is defined in terms of how uniformly a test predicts some outcome variable (e.g., school achievement), for different groups, but in terms of mean group differences, with the test showing a lesser group difference defined as the less biased test. Jensen (1980), for example, has characterized this view (that mean differences between groups are evidence of bias) as the "egalitarian fallacy." However, the more verbally and quantitatively weighted tests are better predictors of school success than nonverbal measures, particularly nonverbal measures that are limited in content, such as the NNAT, which consists only of figural reasoning items (Lohman, 2005). In this sense, comprehensive measures of intelligence are less biased than strictly nonverbal measures of intelligence because they are better at predicting school success for

^{*} George Manos Ed. S., NCSP is the Lead Psychologist for Baker County School Board, Florida, and may be contacted at >gmanos1957@yahoo.com<

members of all groups to which they are applied, and success in a Gifted program in particular. This assumes that the term “gifted” refers to intellectually and/or academically gifted, instead of the Federal definition which includes nonintellectual measures such as creativity, leadership, musical ability, athletic ability, etc. Indeed, the consensus of experts in the field of intelligence testing is that intelligence tests in general are not biased against African-Americans or any other English-speaking native minorities in the United States (Gottfredson, 1997; Herrnstein & Murray, 1994; Maranon & Pueyo, 2000). Furthermore, much evidence now shows that the magnitude of the Black-White IQ difference is a function of the g-loading of the test (Jensen, 1980, 1998; Jensen & Rushton, 2005).

Leaving aside the above theoretical considerations, what about the claim, by Naglieri and Ford (2003), that on the NNAT Black and Hispanic students are as likely to earn high scores as White students? Lohman (2005b) reviewed the evidence for this claim and concluded that Naglieri and Ford’s sample was unrepresentative of the U.S. school population as well as unrepresentative of ethnic groups within that population. Moreover, Lohman noted that the White-Black and White-Hispanic differences reported by Naglieri and Ford (2003) were smaller than the differences reported by Naglieri and Rolling (2000) in a previous analysis of the data involving matched subjects in various demographic categories.

The purpose of the present study was to examine the claim by Naglieri that the NNAT yields equal proportions of White and Black students scoring in the gifted range. An insufficient number of Hispanic students in the total sample precluded consideration of this group.

Method

Participants

Participants were 392 second-grade students ranging in age from 6 years, 11 months to 10 years, 0 months (mean age: 7 years, 10 months), and included 334 White students, 45 Black students, 6 students of mixed Black-White ancestry, 3 Hispanic Students, 2 Asian students, and 2 students designated “other” by the school district. The six students of mixed Black-White ancestry were included in the Black sample. This was done to increase the sample size for the Black sample. Moreover, African-Americans as a whole constitute a mixed-race population, as the average African-American has approximately 20% White ancestry (Chakraborty, Kamboh, Nwankwo, & Ferrell, 1992). 149 of the White students (44.7% of the White sample) and 46 of the Black students (90.2%) were on free or reduced lunch. This statistic was used as a proxy measure for low socioeconomic status.

Statistical Procedures

Mean IQ, standard deviation, and range were computed for the total sample. Mean IQ, standard deviation, range, difference between groups in standard score points, difference between groups in standard deviation units, and a Welch t-test for significance were computed for the Black and White samples. These statistics were also computed for the total sample on free and reduced lunch and the total sample not on free and reduced lunch, for the White sample on free and reduced lunch and the White sample not on free and reduced lunch, and for the White and Black samples, both on free and reduced lunch.

Results

The demographics of the total sample are presented in Table 1.

Table 1
Demographics of Sample

Race/Ethnicity	N	Percentage of Sample
Whites	334	85.1%
Blacks	45	11.5%
Multi-Racial*	6	1.5%
Hispanics	3	0.7%
Asians	2	0.5%

Other	2	0.5%
--------------	----------	-------------

Table 2

Mean IQ, Standard Deviation, and Range for the Total Sample

Total Sample	N	IQ	Standard Deviation	Range
	392	97.0	18.0	50-146

Table 2 summarizes the mean IQ, standard deviation and range for the total sample. The mean IQ of 97.0 is somewhat depressed in relation to the general population mean of 100. This lower score is consistent with the difference in mean IQ between rural and urban populations, in favor of the latter (Jensen, 1981). The greater than 15 point standard deviation for the sample is also consistent with the finding that figural reasoning tests such as the NNAT typically have a greater standard deviation than global measures of intelligence (Jensen, 1980).

Table 3

Mean IQ, Standard Deviation, and Range for the White and Black Samples

	N	IQ	Standard Deviation	Range
Whites	334	98.1	18.0	50-146
Blacks	51	89.0	15.4	61-129
Group Difference – 9.1 (difference of .51 standard deviation units)				
Welch t-test for significance (two-tailed): p = .0003				

The mean IQ, standard deviation and range for the Black and White samples are presented in Table 3. The mean difference between the two samples, although significant ($p < .001$), is less than the typical one standard deviation difference between Whites and Blacks reported in the psychological literature (Eysenck, 1971; Gottfredson, 1997; Jensen, 1980, Jensen & Rushton, 2005; Rushton, 2000; Shuey, 1966; Suzuki & Valencia, 1997) on global measures of intelligence, but is consistent with the smaller difference between Blacks and Whites on some nonverbal measures of intelligence. The smaller standard deviation for the Black sample is also consistent with other intelligence measures, with Black samples typically exhibiting less variability than White samples (Gottfredson, 2005; Jensen, 1980, 1998).

Table 4

Mean IQ, Standard Deviation, and Range for the Total Sample not on Free or Reduced Lunch and the Total Sample on Free or Reduced Lunch

Total sample not on free or reduced lunch	N	IQ	SD	Range
(49.5% of total sample)	194	100.6	16.8	54-146

Total sample on free or reduced lunch

(50.5% of total sample)

198 93.5 18.5 50-146

Group Difference – 7.1 (difference of 0.40 standard deviation units)

Welch t-test for significance (two-tailed): p = .0001

The mean, standard deviation and range of the two samples, one not on free or reduced lunch and the other on free or reduced lunch, are presented in Table 4. The lower mean IQ for the sample on free or reduced lunch (used as a proxy measure for low socioeconomic status) in comparison to the sample not on free or reduced lunch ($p < .001$) is consistent with the well established correlation between socioeconomic status and IQ (White, 1982); however, such a correlation gives no indication of the direction of causality. Studies where genetic factors are held constant, e.g., twin studies involving both monozygotic and dizygotic twins reared in different households of varying socioeconomic status, demonstrate the predominant influence of genes over socioeconomic status in determining adult IQ (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). Furthermore, adopted children, whose IQs by adulthood have a zero correlation with the unrelated individuals with whom they were reared, demonstrate that shared family environments are not a strong determinant of a child's intelligence. (McGue, Bouchard, Iacono, & Lykken, 1993). Table 5 shows the relationship between IQ and socioeconomic status with race held constant ($p < .01$).

Table 5

Mean IQ, Standard Deviation, and Range for the White Sample not on Free or Reduced Lunch and the White Sample on Free or Reduced Lunch

	N	IQ	SD	Range
Whites not on free or reduced lunch (55.3% of white sample)	185	100.6	16.7	54-146
Whites on free or reduced lunch (44.7% of white sample)	149	95.1	19.0	50-146

Group Difference – 5.5 (difference of .31 standard deviation units)
Welch t-test for significance (two-tailed): p = .0059

Table 6 presents the mean IQ, standard deviation, and range for both the White and Black samples on free or reduced lunch status. Matching Blacks and Whites for SES typically reduces the Black/White IQ difference by one-fifth, from 15 to 12 standard score points (Herrnstein & Murray, 1994). With the current sample the difference was reduced from 9.1 to 6.0, a difference of 3.1 standard score points, although this difference was still significant ($p < .04$). Using standard deviation units, the difference was reduced from .51 to .33 standard deviation units, a difference of .18 standard deviation units, an approximately one-third reduction in the difference. However, as previously noted, matching subjects for SES to some degree also controls for genetics, as some proportion of SES differences is genetic (Plomin & Petrill, 1997).

Table 6

Mean IQ, Standard Deviation, and Range for the White and Black Samples, both on Free or Reduced Lunch

Whites on free or reduced lunch (44.6% of white sample)	N	IQ	SD	Range
	149	95.1	19.0	50-146
Blacks on free or reduced lunch (90.2% of black sample)	46	89.1	15.5	61-129

Group difference – 6.0 (difference of 0.33 standard deviation units)
Welch t-test for significance (two-tailed): $p = .0350$

Discussion

Why is the Black/White difference in this study less than the typically reported one standard deviation difference, even before matching black and white students by free and reduced lunch status? This smaller difference is consistent with the findings of Dickens and Flynn (2006), who analyzed data from nine standardization samples for four major intelligence tests. For the period from 1972 to 2002, using data from the various standardizations of the Wechsler Intelligence Scale for Children (WISC-R, WISC-III, and WISC-IV), two standardizations of the Wechsler Adult Intelligence Scale (WAIS-R and WAIS-III), two standardizations of the Stanford-Binet Intelligence Scale (SB-4 and SB-5), and two standardizations of the Armed Forces Qualification Test (AFQT), they report a gain of 4 to 7 standard score points for Blacks in relation to non-Hispanic Whites. Rushton and Jensen (2006) counter that Dickens and Flynn were selective in their analysis, excluding tests showing lesser gains for the time period, such as the Wonderlic Personnel Test, the Woodcock-Johnson Tests of Cognitive Abilities, and the Differential Ability Scale, or tests showing a slight increase in the difference, such as the Kaufman Assessment Battery for Children. Moreover, Rushton and Jensen indicate that the ostensible 4 to 7 point gain is an extrapolation and not based on actual data points, with Rushton and Jensen calculating a mean gain of 3.44 standard score points using actual data points.

Murray (2008), examining three standardization samples of the Woodcock-Johnson for the period 1976 to 1999, reported a reduction in the Black/White difference of approximately 3 standard score points, with this reduction still resulting in a 1.05 standard deviation difference. Another finding, relevant to the present study, is that the Black/White IQ difference increases with age, with average differences of approximately 0.7, 1.0, and 1.2 standard deviation units for early childhood, middle childhood, and late adolescence and early adulthood, respectively (Jensen, 1998). This pattern is consistent with the present study, in which a 0.51 standard deviation difference was manifested, and is also an alternative explanation to the findings of Flynn and Dickens.

In the present study, of the total White sample, 55 students, or 16.5% obtained a cut-score of 115, and for the total Black sample, 1 student, or 2.0% met the cut-score of 115. These are the approximate percentages one would expect given the mean score differences between the two groups. Indeed, using a cut-score of 130, which is the traditional minimum score required for selection into Gifted programs, the disproportionality was even more extreme, with 10 White students (3.0%) obtaining a score of 130 or above and no Black students (0.0%) obtaining such a score. When two groups have different means on any test, the group with the lower mean will be disproportionately overrepresented below a particular cut-score and disproportionately underrepresented above the same cut-score; hence, “the disproportionality problem.” Naglieri’s claim that the NNAT produces equal proportions of students of various ethnicities scoring in the gifted range is not supported by the present study.

Limitations

Some shortcomings of the present study will now be discussed. One problem is possible sampling errors, due to the small sample size for the Black sample. There were also insufficient numbers of Hispanic and Asian subjects to allow for statistically meaningful comparisons. A more refined measure of SES would also have been useful. Another issue is that the g-loading of this test for this particular age level is unknown. All IQ tests are not equally sound psychometrically with respect to reliability, ceiling/floor effects etc, or equally “g-loaded.” Some tests are better measures of general mental ability, while others are “contaminated” by a greater proportion of specific factors. “Inadequate g loading may be a problem, especially in the early grades, because it can be difficult at these ages to measure g adequately with group rather than individually administered IQ tests, but group tests are the only feasible way to test large samples.” (Gottfredson, 2005).

References

- Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science*, 250, 223-228.
- Chakraborty, R., Kamboh, M. I., Nwankwo, M., & Ferrell, R. E. (1992). Caucasian genes in American blacks: New data. *American Journal of Human Genetics*, 50 (1), 145-155.
- Dickens, W. T. and Flynn, J. R. (2006). Black Americans reduce the racial IQ gap. *Psychological Science*, 17 (10), 913-920.
- Eysenck, H. J. (1971). *Race, Intelligence, and Education*. London: Temple Smith.
- Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24 (1), 13-23.
- Gottfredson, L. S. (2005). Implications of cognitive differences for schooling within diverse societies. In: C. L. Frisby & C. R. Reynolds (Eds.): *Comprehensive Handbook of Multicultural School Psychology*, pp. 517-554. New York: Wiley.
- Herrnstein, R. J. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Simon & Schuster.
- Jensen, A. R. (1980). *Bias in Mental Testing*. New York: Free Press.
- Jensen, A. R. (1981). *Straight Talk about Mental Tests*. New York: Free Press.
- Jensen, A. R. (1998). *The g factor*. London: Praeger.
- Jensen, A. R. and Rushton, J. P. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11 (2), 235-294.
- Lohman, D. F. (2005). The role of non-verbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, 49, 111-138.
- Lohman, D. F. (2005b). Review of Naglieri and Ford (2003): Does the Naglieri Nonverbal Ability Test identify equal proportions of high-scoring White, Black, and Hispanic students? *Gifted Child Quarterly*, 49 (1), 19-28.
- Maranon, R. C., & Pueyo, A. A. (2000). The study of human intelligence: A review at the turn of the millennium. *Psychology in Spain*, 4 (1), 167-182.

McGue, M., Bouchard, T. J., Jr., Iacono, W. G., & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In: R. Plomin & G. E. McClearn (Eds.): *Nature, Nurture, and Psychology*, pp. 59-76. Washington D. C.: American Psychological Association.

Murray, C. (2008). The magnitude and components of change in the black-white IQ difference from 1920 to 1991: A birth cohort analysis of the Woodcock-Johnson standardization. *Intelligence*, in press.

Naglieri, J. A. & Rolling, M. E. (2000). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, 12 (3), 328-334.

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.

Plomin, R. & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence*, 24 (1), 53-77.

Rushton, J. P. (2000). *Race, Evolution, and Behavior: A Life History Perspective*, third edition. Port Huron: Charles Darwin Research Institute.

Rushton, J. P. and Jensen, A. R. (2006). The totality of available evidence shows the race IQ gap still remains. *Psychological Science*, 17 (10), 921-922.

Shuey, A. M. (1966). *The Testing of Negro Intelligence*, second edition. New York: Social Science Press.

Suzuki, L. A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence: Educational implications. *American Psychologist*, 52 (10), 1103-1114.

White R. K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91 (3), 461-481.

