Running head: REVIEW OF NAGLIERI AND FORD

Review of Naglieri and Ford (2003):

Does the Naglieri Nonverbal Ability Test

identify equal proportions of high-scoring White, Black, and Hispanic students?

David F. Lohman

The University of Iowa

Draft of December 9, 2003

To appear in *Gifted Child Quarterly*.

Abstract

In a recent article in this journal, Naglieri and Ford (2003) claimed that Black and Hispanic students are as likely to earn high scores on the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) as White students. However, the sample that Naglieri and Ford used was not representative of the U.S. school population as a whole, and was quite unrepresentative of ethnic subgroups within that population. For example, only 5.6% of the children in their sample were from urban school districts, and both Blacks and Hispanics were relatively more likely to be high SES than were Whites. Further, the mean and standard deviation in the sample differed significantly from the Fall NNAT test norms, which were based on the same data. Finally, White-Black and White-Hispanic differences reported by Naglieri and Ford (2003) were smaller than those reported by Naglieri and Ronning (2000a) in a previous analysis of the same data set in which students were first matched on several demographic variables. For these reasons, I argue that the authors' claims are not supported by the data that they present.

Review of Naglieri and Ford (2003): Does the Naglieri Nonverbal Ability Test (NNAT) identify equal proportions of high-scoring White, Black, and Hispanic students?

In a recent article in this journal, Naglieri and Ford (2003) claimed that Black and Hispanic students were as likely to earn high scores on the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) as White students. Because of this, they argued that this test is the preferred measure of ability for identifying gifted children. It is especially to be preferred, they argue, over ability tests that also have verbal and quantitative sections. Verbal and quantitative abilities are developed through schooling, and so it is argued that tests that measure these abilities would be inappropriate for identifying academically gifted minority students.

Strong claims are made for the NNAT. The test is said to be culture fair (Naglieri, 1997b); to show at most small and inconsequential mean differences between minority and White students (Naglieri & Ronning, 2000a); to predict achievement as well as measures of ability that contain both verbal and nonverbal content (Naglieri, 2003b; Naglieri & Ronning, 2000b); and, finally, to identify equal numbers of high-scoring Black, Hispanic, and White students (Naglieri & Ford, 2003). In short, Naglieri claims to have accomplished what many test developers have tried to do since at least the 1930s, but with at most limited success. Surprisingly, he did this by creating test items in the format of the Progressive Matrices Test (Raven, Court, & Raven, 1983), grouping them somewhat differently by test level, and then printing the test stimuli in two colors. How these changes succeed in eliminating group differences in ability that have emerged on every other well-constructed ability and achievement test battery is unclear.

Academics commonly argue about how data should be analyzed. Those who are not skilled in such matters either skip over the details of the analyses or yawn and wait for a verbal explanation of the results. Often this is not an unreasonable thing to do. Arguments about methodology have no beginning and seemingly no end. Here, however, the consequences are more serious than whether one method of analysis is to be preferred over another. Children's

lives are affected by whether one believes the Naglieri and Ford (2003) data and the claims made for the NNAT. When new drugs are proposed, we require that the drug company's claims be vetted by the FDA. Education has no such agency. We police ourselves, sometimes, it seems, to our peril.

Are These Claims Plausible?

Elsewhere, I review the NNAT and challenge the claims that figural reasoning tests are culture fair, that they predict achievement as well as tests that contain verbal and quantitative content, or that they should be used as the primary screening measure to identify students for inclusion in programs for academically gifted students (Lohman, 2003, in press). In this article, I focus on the claim that the NNAT identifies equal proportions of White, Black, and Hispanic students.

First, however, it is necessary to consider the plausibility of this claim. Many competent and dedicated people have tried to develop tests that would give score distributions with the same mean and same variance for different ethnic and cultural groups. Since the earliest days of mental testing, test developers have created hundreds of different ways to estimate human abilities in presumably "culture-free" or "culture-fair" ways. In spite of this effort, no one has yet found a way to eliminate the effects of ethnicity, education, or culture on tests that measure either abstract reasoning abilities or important outcomes of education (Anastasi & Urbina, 1997; Jencks & Phillips, 1998). Not all tests show equal differences between ethnic groups. Unsurprisingly, differences between native and nonnative speakers are largest on verbal tests that require detailed knowledge of the syntax and structure of the English language. Although this does not mean that the test is biased, it can mean that proper interpretation of scores requires comparison to a group of children with similar language experiences as well as to the same age or grade peers (Lohman, in press).

Black-White differences, on the other hand, seem to vary primarily as a function of the general factor loading of the test (Jensen, 1998). Differences are often largest on nonverbal,

figural tests that require spatial thinking, and smallest on tests that require speeded comparisons of stimuli. Indeed, Black students often score better on tests that use verbal or quantitative content than on tests with spatial/figural content, especially tests that use spatial transformations such as mental rotation (Jensen, 1998). Spatial visualization items constitute 5 of the 38 items at level C of the NNAT, 14 at level D, 19 at level E, 18 at level F, and fully 24 of the 38 items at level G. That there are no differences between the proportion of high-scoring Blacks and Whites on a test that contains such a preponderance of such items is surprising.

One can reduce the size of any ethnic difference by removing sources of difficulty that, for one reason or another, happen to be correlated with ethnicity. Most well-constructed ability and achievement tests go to great lengths to detect and remove items that contain irrelevant sources of difficulty. But one can also reduce group differences by building tests or constructing score composites that give greater weight to abilities that show smaller group differences. For example, Jensen (1984) argued that the main reason the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983) showed Black-White differences that were about half as large as those that had been observed on other individually administered intelligence tests was that it was a poorer measure of g. Shorter, less reliable tests will also show smaller group differences than longer, more reliable tests.

Most well-constructed ability and achievement tests, however, show substantial differences between the scores of White and Black students, especially in the proportion of each group who receive high scores on the test. For example, Hedges and Nowell (1998) found that the proportion of White students who scored in the highest five percent of the score distributions on a wide variety of achievement and ability tests (including nonverbal reasoning and spatial tests) to be 10 to 20 times the proportion of Blacks who scored in the same range (see also Koretz, Lynch, & Lynch, 2000). Indeed, the NNAT uses a well known and much studied item format introduced many years ago by J. C. Raven. Scores on Raven's Progressive Matrices tests (Raven et al., 1983) show large group differences when individuals are grouped by ethnicity,

SES, education, or age cohort. For example, Raven (2000) reported 1986 norms on the Standard Progressive Matrices Test for adolescents in one large U.S. school district. The median score for Hispanics fell at the 23rd percentile of the White distribution. The median score for Blacks fell at the 17th percentile of the White distribution. These correspond to effect sizes of .73 *SD* (for the Hispanic-White comparison) and .94 *SD* (for the Black-White comparison) in normally distributed scores.

What alterations did Naglieri make to the content and format of progressive matrix items that would eliminate these large and consistently observed differences between ethnic groups? Naglieri's adaptation of Raven's test involved two major changes: (a) systematically constructing items to emphasize pattern completion, analogical reasoning, series completion, or spatial visualization, and then blocking items so that different levels of the test contained different mixes of items (and thus measure somewhat different abilities), and (b) printing items in two colors rather than in many colors, as in Raven's Colored Progressive Matrices or in black-and-white, as in the Standard and Advanced Progressive Matrices tests. I know of no research that would lead one to expect that either of these manipulations would reduce, much less eliminate, ethnic differences.

Thus, I find Naglieri and Ford's (2003) claims implausible because they run counter to what other investigators have found using similar tests and procedures.[1]

A Closer Look at the Data

Dependable inferences about populations require representative samples. Thus, the first step in understanding the Naglieri and Ford (2003) data is to examine how the sample was collected and whether it is representative of the several populations to which inferences are made.

The data used in the Naglieri and Ford (2003) study were part of a larger data set collected as part of the NNAT test standardization. In the NNAT manual, Naglieri (1997b) reports that 22,600 students were tested in the fall of 1995 and another 67,000 in the spring of 1996.[2] Why the larger Spring data set or the combined Spring and Fall data sets were not used

for the analyses is not explained. Only 30% of schools and districts that were invited to participate in the Fall testing agreed to participate in the standardization study. The sample was thus not representative of the planned national sample. Data were then weighted in an effort to make the obtained sample more representative of the intended population. This was done by deleting or duplicating student records until the sample better approximated the nation in terms of demographic characteristics. How much the data had to be changed to make this happen is not reported. Only the percentages of students in each demographic category for the final, weighted sample are reported.[3]

Of the many things that one might want to know about a sample, the most important is the number of cases. Naglieri and Ford (2003) reported that "the sample included 20,270 children from the NNAT standardization sample tested during the fall of 1995" (p. 157). Table 1 in their paper shows that 14,141 of these students were White, 2,863 were Black, and 1,991 were Hispanic. However, these numbers total 18,995, not 20,270. Their Table 1 also shows the breakdowns of these 18,995 cases by gender, region of the country, urbanicity, and SES. For example, of the 14,141 White students, 7,090 were male and 7,088 were female. But if one adds the last two numbers, the total is 14,178, not 14,141. Similarly, adding the number of Whites across the four regions gives 14,180, across the three urbanicity or five SES categories gives 12,279. Perhaps the reported sample size refers the number of cases with scores on any variable, in which case the total number of White students should be at least as large as the largest of these subtotals (i.e., 14,180). Here, this is not the case. Nor can the reported sample size be the number of cases with complete data on all variables, since that *N* would be no greater than 12,279. Thus, it is unclear whether the number of Whites in the sample was 12,279, 14,141, 14,178, 14,180 or some larger number that would, when combined with the number of Blacks and Hispanics, add to the reported total sample size of 20,270.

The entries for Blacks and Hispanic students show similar inconsistencies. The authors report that 2,883 Blacks and 1,991 Hispanics were included in the study. But adding the number

of males and females gives 2,865 Blacks and 1,997 Hispanics. Subtotals for breakdowns by region, urbanicity, and SES also vary. Briefly, the number of Blacks ranges from 2,737 to 2,865 and the number of Hispanics from 1,934 to 2,002. Neither the smallest nor the largest of these numbers coincide with the reported sample sizes for each group.

Leaving aside the question of sample size, the next important issue is the representativeness of the sample. There are two concerns here. First, was the weighted Fall sample representative of the U.S. population? For example, if 16.1% of the U.S. population is Black, was approximately 16% of the sample Black? The target national percentages for each demographic category used in the sampling plan and the observed percentages for the weighted fall sample are shown in the first two columns of Table 1. The differences between the target and observed percentages are shown in the third column. The most notable problems are the substantial underrepresentation of students from urban school districts and the corresponding overrepresentation of students from rural districts. Only 5.6% of the students were from urban school districts. This is markedly below the population value of 26.8%. This says that the overall norms for the test do not really describe the U.S. school population. More important, however, is the fact that urban school districts have high concentrations of minority students who score poorly on ability and achievement tests. Indeed, about 80% of Blacks who scored in the bottom quintile on NAEP during the 1990s lived in the South or in large urban school districts in the Midwest and Northeast. About 70% of Hispanics in the lowest quartile lived in urban school districts in the South or West (Flanagan & Grissmer, 2001). It is therefore quite likely that the minority students who were included in the sample were not a representative group.

This leads to the second point. Was each of the subsamples detailed in the Naglieri and Ford (2003) article (Whites, Blacks, Hispanics) representative of its respective population? This is the question that must be addressed when comparing within-ethnic group score distributions. It is a much more difficult standard to achieve than simply having a requisite percentage of each ethnic group in the overall sample. For example, the percentage of Blacks who live in the South

is much greater than the percentage of Blacks who live in other regions of the U.S. Thus, a representative sample of U.S. Blacks should have a greater percentage of students from the South than would a representative sample of all U.S. students. This means that even if the numbers for the national sample in column 2 perfectly matched the numbers in column 1, it is highly unlikely that the demographic characteristics for various subgroups would match their national profiles.

The entries in Table 2 confirm this suspicion. These show the target percentages for these demographic variables for each of the three ethnic groups. Most Blacks in the Naglieri and Ford sample lived in the West (40% in the NNAT sample versus 17% in the U.S. Census) and attended suburban schools (54.1% in the NNAT sample versus 31.7 % in the U.S. Census). Census data, on the other hand, show that Blacks are much more likely to live in the South (38.8% in the U.S. Census versus 19.4% in the NNAT sample) and in large metropolitan areas (54.9% in the U.S. Census versus 11.0% in the NNAT sample). Socioeconomic status (defined by an unspecified combination of income and education) now shows a bizarre bimodal distribution. In the sample Naglieri and Ford used, only 8.4% of Blacks were classified as middle SES. More (19.5%) were high-middle SES; and slightly more than one-quarter (25.2%) were high SES. Similarly, only 3% of Hispanics were middle SES; 6.2% were high-middle SES; and 19.5% were high SES. In other words, both Blacks and Hispanics were more likely to be high SES than middle SES, and even more likely than Whites to be high SES! These trends run completely counter to the 2000 U.S. Census data, which show an increasing disparity in the percentage of White and minority families as one moves up the scales for income and education. High SES students are the most likely to show high scores on tests of all sorts. Thus, these data should not be used for making generalizations about the performance of U.S. school children in general, and are quite inappropriate for making generalizations about the distributions of scores for Black and Hispanic students.

Although the data used in this study are not representative of the nation, they should be representative of the NNAT Fall norms. The 18,995 students included in the Naglieri and Ford

(2003) sample constitute the majority of the students who participated in the NNAT Fall norm sample. The Fall test norms were based on these students' test scores. Thus, the mean and standard deviation of the score distributions should approximate the Fall NNAT Norms. The NNAT is scaled so that it has a population mean of 100 and standard deviation of 15.0. Naglieri and Ford reported means and standard deviations (*SD*) for 14,141 Whites, 2,863 Blacks, and 1,991 Hispanics. Scores for the remaining 1,275 students in their sample from other ethnic backgrounds (notably Asian Americans) were not reported.[4] Oddly, the means for all three of the reported groups were less than 100. A little algebra reveals that the *mean* score for the excluded students would have to be 120.7 to bring the population mean up to 100. Since it is unlikely that the typical excluded student obtained a score of 120.7, it is reasonable to conclude that the mean score Naglieri and Ford reported differs from the NNAT Fall norm mean that was based on the same data.

The standard deviations that Naglieri and Ford (2003) reported also differ from the Fall norm value of 15.0. Oddly, the *SD*s for all three groups were greater than 15.0. *SD*s were 16.7, 16.8, and 17.5 for White, Hispanic, and Black students, respectively. A simple *F* test on the variances that correspond to these three *SD*s shows all three *SD*s to be significantly greater than the population *SD* of 15.0. Pooling all three estimates gives an estimated population *SD* of 16.9, which is considerably above the population value of 15.0. The *SD* is critical to the conclusions Naglieri and Ford (2003) offer because inferences about the percentages of students at the tails of the distributions depend on the relative *SD* of each group. The fact that the SD of the score distribution of Black students is greater than the score distribution for White students is puzzling. This is puzzling because most national surveys find that the variability of the score distribution for Blacks is considerably smaller than the variability of the score distribution for Whites (Hedges & Nowell, 1998).[5] Here the Black distribution was considerably more variable. Perhaps the same factor that produced the unusual distributions for SES also inflated the variances.

Would including scores for the excluded students bring the *SD* back to 15.0? Even if all of these excluded students had the same score (i.e., *SD* for this group = 0), the population *SD* would be 16.3, not 15.0. In other words, the Naglieri and Ford Fall normative data do not generalize to the NNAT Fall norms, much less to the larger population or to ethnic groups within that population.

Finally, even if the Naglieri and Ford (2003) data are not representative of the national samples, or of the NNAT Fall norms, one would expect that different analyses on the same data set would yield consistent results. In an earlier analysis of this same data set, Naglieri and Ronning (2000a) investigated ethnic differences in scores on the NNAT and the Stanford Achievement Test (9th ed., 1995) for samples of students who were matched on SES, geographic region, urbanicity, school level (elementary, middle, or high school), and sex. Mean differences in NNAT scores for matched groups of White and Black, White and Hispanic, and White and Asian American students from this study are shown in the first row of Table 3. Consistent with many other studies, controlling for the effects of these demographic variables gave small group differences on both the NNAT and the Stanford Achievement Test.[6]

Naglieri and Ford (2003) used the same data set, but did not control for these factors. Therefore, one would expect the differences between White and minority students to be substantially larger. The mean differences between groups of unmatched students in the Naglieri and Ford (2003) study are shown in the second row of Table 3. Unexpectedly, they are actually smaller than the mean differences for the matched cases in the Naglieri and Ronning (2000a) analyses of the same data set. It is difficult envision a plausible reason why controlling for SES, region of the country, urbanicity, level in school, and sex would increase the magnitude of group differences on any test of ability or achievement.

Summary

The claim that the NNAT identifies equal proportions of high-scoring White, Black, and Hispanic students is, on its face, exceedingly implausible. For normally distributed scores, even

small differences between the mean scores of groups translate into much larger differences in the relative proportions of each group at the tails of the distribution. These effects are magnified if the lower scoring group also has a less variable score distribution, which is the repeated observation in large national surveys that compare the score distributions for Whites and Blacks.[7] A second look at the Naglieri and Ford data shows that the sample is not representative of the U.S. school population as a whole, or of ethnic subgroups within that population. Urban school children were markedly underrepresented. Both Blacks and Hispanics were relatively more likely to be high SES than were Whites in this data set. Even more problematic is the fact these data do not reproduce the Fall NNAT test norms, which were based on them. In particular, the mean is significantly lower, and the variance significantly greater than the published norms that use the same data. Finally, mean differences between Whites and Blacks and between Whites and Hispanics reported by Naglieri and Ford (2003) are actually smaller than the corresponding mean differences obtained in other analyses of the same data in which students were first matched on SES, region of the country, urbanicity, school level, and sex.

One can achieve the goal of identifying more academically talented minority students. Elsewhere I offer suggestions for doing this in a way that not only identifies more minority students, but, more importantly, more students who are more likely to attain academic excellence (Lohman, in press). Although nonverbal, figural reasoning tests do have a role to play in this process, it is always as an ancillary measure, not as the primary means of identification. This is because although nonverbal tests test may appear to reduce bias, when used alone they actually increase it by failing to identify those minority or majority students who either currently exhibit academic excellence or who are most likely to profit from educational enrichment.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Flanagan, A., & Grissmer, D. (2001, February). *The role of federal resources in closing the achievement gaps of minority and disadvantaged students.* Paper presented at the Brookings conference on the Black-White test score gap.

Hedges, L. V., & Nowell, A. (1998). Black-White test score convergence since 1965. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 149-181). Washington, DC: Brookings Institution Press.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Jensen, A. R. (1984). The Black-White difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18,* 377-408.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kaufman, A. F., & Kaufman, N. L. (1983). *The Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

Koretz, D., Lynch, P. S., & Lynch, C. A. (2000). The impact of score differences on the admission of minority students: An illustration. *Statements of the National Board on Educational Testing and Public Policy, 1* (5).

Lohman, D. F. (2003). A comparison of the Naglieri Nonverbal Ability Test (NNAT) and Form 6 of the Cognitive Abilities Test (CogAT). Retrieved December 12, 2003, from The University of Iowa, College of Education Web site: *http://faculty.education.uiowa.edu/dlohman*.

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

Naglieri, J. A. (1997a). *Naglieri Nonverbal Ability Test.* San Antonio, TX: Harcourt Brace.

Naglieri, J. A. (1997b). *Naglieri Nonverbal Ability Test: Multilevel technical manual.* San Antonio, TX: Harcourt Brace.

Naglieri, J. A. (2003a, November). Assessment of diverse populations with the *Naglieri Nonverbal Ability Tests*. Paper presented at the meeting of the National Association for Gifted Children, Indianapolis, IN.

Naglieri, J. A. (2003b). Fair assessment of minority children with Verbal/Achievement reduced intelligence tests. Retrieved December 12, 2003, from The George Mason University web site: http://ccd.gmu.edu/Naglieri/Fair%20assessment%20NASP%20%2703.htm

Naglieri, J. A. (2003c). Testing diverse populations with the Naglieri Nonverbal Ability Test (NNAT). Retrieved December 12, 2003, from http://www.mypsychologist.com/nnat_min.htm.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System.* Itasca, IL: Riverside.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47,* 155-160.

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

Naglieri, J. A., & Ronning, M. E. (2000b). The relationship between general ability using the Naglieri Nonverbal Ability Test (NNAT) and Stanford Achievement Test (SAT) reading achievement. *Journal of Psychoeducational Assessment, 18,* 230-239.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41,* 1-48.

Raven, J. C., Court, J. H., & Raven, J. (1983). *Manual for Raven's Progressive Matrices and Vocabulary Scales, section 4: Advanced Progressive Matrices, sets I and II.* London: H. K. Lewis.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition.* Itasca: Riverside.

Author Note

David F. Lohman, College of Education.

I am grateful to colleagues and friends for helpful comments on earlier drafts of this manuscript. I also thank Patricia Martin for assistance in preparing the manuscript.

[1] Naglieri and Ford (2003) cite the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983) and the Cognitive Assessment System (CAS; Naglieri & Das, 1997) as the only individually administered ability tests that, like the NNAT, identify similar percentages of White, Black, and Hispanic students. The CAS Interpretive Handbook does not report means and standard deviations for different ethnic groups, and so the claim could not be verified. For the K-ABC, the mean Black-White difference is reported to be 3.0 points on the Sequential Processing Scale, 8.5 points on the Simultaneous Processing Scale, and 7.0 points on the Mental Processing Composite. Although this is smaller than the 10-12 point difference on tests such as the Stanford Binet IV (Thorndike, Hagen, & Sattler, 1986), it is certainly not zero. Jensen (1984) argued that the Black-White difference is smaller on the K-ABC because the K-ABC "yields a more diluted and less valid measure of g than do other tests" (p. 377). Previous analyses of the NNAT data by Naglieri and Ronning (2000a) also do not support the assertion that the NNAT shows only small differences between ethnic groups. These analyses looked only at differences in groups that were first matched on SES, region, urbanicity, level in school, and gender. By this standard, the Stanford Achievement Tests also showed small group differences. See also footnote 6.

[2] The size of the NNAT standardization sample is reported differently in different places. The Technical Manual (Naglieri, 1997b) reported that 22,600 students were tested in the fall of 1995 and 67,000 in the spring of 1996. However, Naglieri and Ronning (2000a) claimed that the Fall sample of 22,620 students were "the portion of the 68,000 children used to standardize the NNAT … that was also administered measures of math and reading" (p. 329). Naglieri (2003c) also reported that the NNAT "was standardized on a sample of 68,000 children" (p. 1). However, Naglieri (2003a) claimed that "the NNAT was standardized on a sample of 89,000 children" (p. 1). All published secondary analyses have used only portions of the 22,600 students sampled in the fall standardization.

[3] Most test developers report demographic characteristics of both the initial, unweighted sample and of the final, weighted sample. This allows the user to see how much the data had to be weighted.

[4] In these analyses, I assume that the sample size was 20,270, as reported in Naglieri and Ford (2003). If the sample size was 22,620 (Naglieri & Ronning, 2000a), 22,600 (Naglieri, 1997b), or 22,490 (Naglieri & Ronning, 2000a, Table 1), then somewhat different means and variances are obtained. In no case, though, does the mean or *SD* approximate the reported population mean and *SD*, even under the surely untenable assumption that all missing cases obtained the same score.

[5] The median Black variance/White variance ratio was .76 across six large national surveys.

[6] After controlling for demographic variables, Black students actually performed better on the Reading and Math achievement tests than on the NNAT. Asian-Americans performed better on the Math achievement test than on the NNAT. These results were ignored in the interpretation of the study.

[7] See Koretz, Lynch, & Lynch (2000) for examples of the consequences of a mean difference of .8 *SD* and a variance ratio of .81 (higher scoring/lower scoring group) for the composition of the group that is selected when two different cut scores are used.

Table 1

Demographic Characteristics (in Percent) of the Fall NNAT Sample used in Naglieri and Ford (2003)

	NNAT	U.S.	Difference
Ethnicity			
White	69.4	66.6	2.8
Black	14.1	16.1	-2.0
Hispanic	10.5	12.7	-2.2
Asian	2.9	3.6	-0.7
American Indian	1.4	1.1	0.3
Region			
Northeast	20.6	19.6	1.0
Midwest	24.2	23.8	0.4
Southeast	20.2	24.1	-3.9
West	34.9	32.4	2.5
Urbanicity			
Urban	5.6	26.8	**-21.2**
Suburban	48.3	48.0	0.3
Rural	46.1	25.2	**20.9**
SES			
Low	20.0	19.6	0.4
Low middle	23.4	21.4	2.0
Middle	15.4	20.8	-5.4
High middle	19.9	17.8	2.1
High	21.2	20.3	0.9

Note. U. S. Population values from NCES (1993-1994 update) for 1990 (from Naglieri , 1997b).

Table 2

Demographic Characteristics (in Percent) of the NNAT Sample used in Naglieri and Ford (2003)

	White			Black			Hispanic		
	NNAT	U.S.	Diff.	NNAT	U.S.	Diff.	NNAT	U.S.	Diff.
Region									
Northeast	15.7	23.6	-7.9	23.7	24.6	-0.9	9.6	14.8	-5.2
Midwest	32.6	28.5	4.1	16.9	19.6	-2.7	6.8	8.0	-1.2
Southeast	24.4	21.2	3.2	19.4	38.8	**-19.4**	11.4	8.8	2.6
West	27.3	26.7	0.6	40.0	17.0	**23.0**	72.1	68.5	3.6
Urbanicity									
Urban	3.3	29.5	**-26.2**	11.0	54.9	**-43.9**	31.2	46.4	**-15.2**
Suburban	44.6	50.3	-5.7	56.1	31.7	24.4	42.8	45.1	-2.3
Rural	52.1	20.2	**31.9**	32.8	13.3	**19.5**	26.0	8.5	**17.5**
SES									
Low	19.2	—	—	20.8	—	—	42.0	—	—
Low middle	20.1	—	—	26.2	—	—	29.3	—	—
Middle	20.4	—	—	**8.4**	—	—	**3.0**	—	—
High middle	23.7	—	—	**19.5**	—	—	**6.2**	—	—
High	16.6	—	—	**25.2**	—	—	**19.5**	—	—

Note. Large discrepancies in bold. Region is defined by NCES using the definition of the Office of Business Economics, U.S. Department of Commerce (rather than the Census bureau). Urbanicity is defined as in the U.S. Census. SES was a composite of median family income in the community and the percent of adults with high school diplomas. However, exactly how each was weighted was not specified in Naglieri (1997b). Dashes indicate that within-group estimates of the target national percentages for the same SES categories could not be derived for White, Black, and Hispanic subsamples. Other estimates of SES,

however, show a decrease in the percentage of minority families in each category as one moves up the SES scale. Diff. = NNAT percent – U.S. percent.

Table 3

Mean White-Black and White-Hispanic Differences in the Naglieri and Ronning (2000a) and the Naglieri and Ford (2003) Analyses of the Fall, 1997, NNAT Data.

Study	White-Black	White-Hispanic
Naglieri and Ronning (2000a)	4.2	2.8
Naglieri and Ford (2003)	3.2	2.0