



The Journal of Sociology & Social Welfare

Volume 29
Issue 3 *September*

Article 6

September 2002

Differential Test Performance in the American Educational System: The Impact of Race and Gender

Stephen J. Finch
Stony Brook University

Harvey A. Farberman
Stony Brook University

Jordan Neus
Stony Brook University

Richard E. Adams
Mt. Sinai School of Medicine

Deirdre Price-Baker
Stony Brook University

Follow this and additional works at: <https://scholarworks.wmich.edu/jssw>

 Part of the [Gender and Sexuality Commons](#), [Race and Ethnicity Commons](#), and the [Social Work Commons](#)

Recommended Citation

Finch, Stephen J.; Farberman, Harvey A.; Neus, Jordan; Adams, Richard E.; and Price-Baker, Deirdre (2002) "Differential Test Performance in the American Educational System: The Impact of Race and Gender," *The Journal of Sociology & Social Welfare*: Vol. 29 : Iss. 3 , Article 6.

Available at: <https://scholarworks.wmich.edu/jssw/vol29/iss3/6>

This Article is brought to you for free and open access by the Social Work at ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.



Differential Test Performance in the American Educational System: The Impact of Race and Gender

STEPHEN J. FINCH

Stony Brook University
Department of Applied Mathematics and Statistics

HARVEY A. FARBERMAN

Stony Brook University
School of Social Welfare

JORDAN NEUS

Stony Brook University
Department of Community and Preventive Medicine

RICHARD E. ADAMS

The Mt. Sinai School of Medicine
Department of Community and Preventive Medicine

DEIRDRE PRICE-BAKER

Stony Brook University
Department of Applied Mathematics and Statistics

Contrary to Herrnstein and Murray (1994) who claim that racial groups have different cognitive endowments and that these best explain differential test score achievements, our regression analyses document that there is less improvement in test scores per year of education for African-Americans and women. That is, the observed group test score differences do not appear to be due to racial cognitive differences but rather to other factors associated with group-linked experiences in the educational system. We found that 666 of the subjects in the Herrnstein-Murray database had actual IQ scores derived from school records. Using these as independent controls

for IQ, we document that each of the test components that were the basis of the Herrnstein-Murray "IQ" scores was significantly associated with education level ($p < .001$). Consequently, their IQ score appears to be an education—related measure rather than an IQ test, and thus challenges the validity of their analysis.

Introduction

Over the last 30 years, the use of social science data to support and evaluate social policy initiatives has become routine (Massy and Denton 1993; Wilson 1987; 1996). Proponents and opponents of welfare reform (Bane and Ellwood 1994; Katz 1997; Mead 1993), for example, have employed data to support alternative policies on welfare use and job training. It is in the area of education, however, that the interplay of research and social policy has generated the most debate. Here, social science data has been used to validate reform efforts in the teaching of mathematics and reading skills, and in support of (or opposition to) affirmative action (Arons 1997; Coleman 1989; Fischer *et al.* 1996; Herrnstein and Murray 1994; Kane 1998).

It is, perhaps, the issue of affirmative action that has generated the most controversy. One side asserts that affirmative action is a waste of resources because environmental interventions cannot overcome the markedly inferior cognitive endowments of certain ethnic and racial groups (Herrnstein and Murray, 1994), while the other side asserts that non-hereditary factors better explain the differential educational performance of ethnic/racial groups (Jencks and Phillips, 1998a).

Our aim here is to examine empirically the performance gap among groups and to document that one specific non-heredity factor (*i.e.*, differential association of test score improvement with education level) offers an alternative explanation of the difference.

Genetic differences have been advanced to account for the differential educational performance of African-American and White students (Jensen 1969; 1985; Herrnstein and Murray 1994). In particular, Herrnstein and Murray (1994) have analyzed Armed Forces Qualification Test (AFQT) data from the National Longitudinal Survey of Youth (NLSY) and have concluded that "intel-

ligence" is distributed unequally among various ethnic groups. The lower academic achievement of African-American students, according to their analyses, is owing to heredity. Indeed, they argue that there will be diminishing returns for African-Americans from the impact of additional years of education on earnings or intelligence. Therefore, social policy efforts to improve African-American students' academic achievement through more years of education or affirmative action are a waste of limited educational resources.

The social policy implications of the Herrnstein and Murray (1994) argument also apply to differences in educational performance across gender or social class. According to their analyses, women who score in the bottom 5 percent on the AFQT are inferior intellectually and are at risk for remaining unmarried, having children out of wedlock, and living in poverty. The main policy recommendation for reducing the poor economic prospects of "dull" women is that they marry and stay married (Herrnstein and Murray, 1994). As with their stance toward minorities, Herrnstein and Murray see little benefit in either spending more money on education, or on altering policies that might reduce the educational and job mobility barriers that women face.

A number of scholars have reviewed the *The Bell Curve* and have questioned whether its "IQ score," and the underlying cognitive endowment it purports to measure, actually has the explanatory power reported by Herrnstein and Murray (1994). (See: Devlin *et al.* 1995a and 1995b; Goldberger and Manski 1995; Gould 1994; Hauser 1995; Heckman 1995; Kohn 1996; Mutaner *et al.* 1996). Furthermore, Korenman and Winship (1995) and Fischer *et al.* (1996) have concluded that, family background may be at least as important as the Herrnstein-Murray IQ score in determining social and economic success of a youth in adulthood. Currie and Thomas (1995) have found that, maternal education and income are important determinants of a child's intelligence test scores.

Turning from family background variables to school and community variables, Coleman *et al.* (1966) have examined the educational environment of urban versus suburban schools and have concluded that school resources *per se* appeared to have little effect on children's performance outcomes. Indeed, subsequent studies have shown that, contrary to popular opinion, and,

despite wide disparities among some inner city school districts and some suburban school districts, on average, there seems to be little difference between urban and suburban schools in either per pupil spending or the quality of facilities (Jencks and Phillips 1998a). Indeed, what seems to be placing White men ahead of African-American men with respect to high school graduation rates, college admission rates, and scores on standardized tests is the fact that urban households typically have higher rates of family poverty and lower levels of educational preparedness of children. As Jencks and Phillips (1998a; 1998b) have suggested, the disparities that still exist between and among groups is a function of how well schools do (or do not) prepare students from such circumstances for jobs in modern, technologically advanced economies.

Specifically, teachers in urban schools were—and are—more likely to have lower expectations of African-American students. These expectations may result in a “Pygmalion effect,” (Rosenthal and Jacobson 1968) such that African-American students come to believe in their academic inferiority and work down to lower expectations and standards. (Ferguson 1998a). In addition, inner city schools were—and are—more likely to have a greater percentage of students with emotional and behavioral problems than suburban schools. These types of students require greater expenditures, thus leaving fewer resources to meet the educational needs of mainstream or above average students (Jencks and Phillips 1998b). Finally, teachers at predominantly African-American schools tend to score lower on standardized tests than teachers at mainly White schools (Ferguson 1998b). For these reasons, some argue that African-American students have had a different, and inferior, educational experience compared to Whites. In other words, American K through 12 public education has not taught academic skills to African-American students as effectively as it has taught them to White students (Phillips, Crouse, and Ralph 1998).

The specific questions that we seek to answer here are: (1) Do students with similar cognitive abilities improve on the AFQT as their years of education increase? (2) Do these improvements go up at the same rate for both African-American and White students? (3) Do they go up at the same rate for male and female

students? (4) If not, to what extent is there differential test score improvement between African-American and White students and male relative to female students?

To answer these questions, our analysis adds the six components of the AFQT *not* considered in Herrnstein's and Murray's work; namely, General Science, Numerical Operations, Coding Speed, Auto and Shop, Mechanical Comprehension, and Electronics. We assume that, these AFQT components assess the acquisition of skills needed to succeed in technologically advanced economies; that is, they measure commercially valuable intellectual skills—for example, the understanding of electronics and mechanics. Specifically, we examine the comparative performances of African-Americans, Whites, men, and women. In particular, we consider a specification of a (regression) analysis that permits a test of whether there is differential test score improvement in per year of education for women and minorities. Further, we control for IQ using a small subset of subjects from the database. Our reasoning is that, **if AFQT scores differ among groups after controlling for both innate ability (IQ score) and other relevant factors**, then differential test performance is a result of environmental factors. Therefore, explanations for any widening test performance gap should be sought among non-heredity factors associated with the educational experience.

Sample

The National Longitudinal Survey of Youth (NSLY) is a nationally representative sample of 12,686 young men and women who were between the ages of 14 and 21 on January 1, 1979, when the study began. Interviewers from the National Opinion Research Council conduct face-to-face interviews with respondents annually. In the 1980-wave of the data collection, nearly 94 percent (N=11,914) of the sample answered questions from the AFQT.

The NLSY used the family as its unit of sampling. The sample consisted of all members of a family between the ages of 14 and 21 who gave consent. As a result, there were a number of siblings and partners in the sample and the residual terms of siblings' intelligence scores may be expected to be high. Herrnstein and Murray (1994, p. 12) reported that the correlation of IQ scores for

monozygotic twins reared apart was 0.78. From basic principles of statistical genetics (Falconer and Mackay, 1996), the residual sibling genetic correlation consistent with this report is 0.39.

Therefore, in the analyses reported in Table 1, we used the oldest sibling with complete information on the variables considered in order to remove the violation of the independence assumption that results from the correlation of test score residuals of siblings within a family. This reduced the number of subjects in the multiple regression analyses to 3,712.

Measures

The AFQT is a multiple-choice test created by the Department of Defense basically to determine two things: (1) whether (or not) to accept a civilian as a recruit, and (2) how best to assign enlisted personnel to military occupations. The questions in the test are scored in ten component parts (that is, Armed Services Vocational Aptitude Battery, abbreviated to ASVAB). These components (General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed, Auto and Shop, Mathematics, Mechanical Comprehension, and Electronics) are excellent measures of skill sets necessary for success in employment and economic mobility.

The validity of the AFQT as a standardized measure of academic skills and as a predictor of economic success has been well established by three decades of research efforts. Jencks and Philips (1998a, 1998b), in their review of research relating AFQT scores to earnings, have argued that reducing African-American-White differences in scores is one of the surest ways to improve upward mobility for African-Americans and for reducing many of the consequences associated with living in poverty. Cutright (1974), for example, has studied 1964 earnings of men in their thirties who had taken the AFQT between 1949 and 1953 and found that, among men with AFQT scores above the national average, African-American men earned 35 percent less than White men. In the 1990's, however, Jencks and Phillips (1998b) found that, as AFQT scores went up, the African-American-White earnings gap narrowed. African-Americans who were above average in 1980 on the AFQT earned 96 percent of the White average by the mid-1990s.

Moreover, Korenman and Winship (1995), Fischer *et al.* (1996), and others have shown that AFQT scores associate closely with years of schooling, types of classes taken (e.g., mathematics), motivation to learn, quality of school's academic environment, performance on standardized tests, and relationships with teachers. They also have shown that these associations occur *regardless* of the students' innate cognitive abilities (Fischer *et al.* 1996; Grassmmer, Flanagan, and Williamson 1998). These authors, along with Phillips *et al.* (1998), have documented that social structural and interpersonal factors such as family background, peers, and community context also affect AFQT scores.

Dependent Variables

We have employed the standardized AFQT score following computation rules used by the military. In addition, we also have analyzed the effect of the independent variables on the ten standardized components of the AFQT. We have analyzed the variables using the SAS statistical package. Following Herrnstein and Murray (1994), we have employed the "standard score" measures of the AFQT and ASVAB components (*i.e.*, the mean of the scores was zero, and their standard deviation was one). Higher scores represent higher levels of skill acquisition or competence. However, we did *not* replicate Herrnstein's and Murray's (1994) scoring of the AFQT but chose to follow the scoring procedure actually reported in the NLSY data. More importantly, we did *not* follow Herrnstein and Murray (1994, p. 569) in using weighted regressions because the *NLSY User's Handbook* (Center for Human Resource Research 1994, p. 463) recommended *against* the use of weights in regression analysis. The exact statement from the *NLSY* guide is:

A common question is whether one should use the provided weights to perform weighted least squares when doing regression analysis. Such a course of action may not lead to correct estimates. If particular groups follow significantly different regression specifications, the preferred method of analysis is to estimate a separate regression for each group or to use dummy (or indicator) variables to specify group membership.

Accordingly, we used indicator variables for the relevant groups.

Independent Variables

We included five background variables: gender, race, age in 1980, the household income of respondent's parents in 1980, and the respondent's education level in 1980. The indicator variable FEMALE was coded (1) for female respondents and (0) for males. Race was coded into two indicator variables, with BLACK coded (1), if the subject self-reported as African-American, and HISP as (1) if the subject self-reported as Hispanic. Whites were the excluded group. Respondent's age (AGE80) was coded to the nearest year. Income variable used units of \$10,000 (INC80), and education was coded to the last year of education completed in 1980 (ED_LVL80).

The NLSY data set over-sampled subjects who were poor, African-American, Hispanic, or in the military (Center for Human Resource Research 1994). That is, there were greater numbers of these groups than would have occurred with purely random probability sampling of subjects. Following the recommendation quoted above, we coded the indicator variable SUPP such that (1) reflects that the subject's family was included as part of the supplemental sample of poor subjects and the value (0) otherwise. Similarly, we coded the indicator variable MIL (1) if the subject's family was included as part of the military sample.

The NLSY also obtained IQ scores from the high school transcripts of a sub-sample of subjects. There were a number of forms of IQ test in use, and a number of subjects had IQ test results reported for two or more forms. To reduce the variability due to the form of the IQ test used, we used the result for the form that was the more commonly reported in the subsample, when a subject had results reported from two or more forms. Thus, there were 666 subjects that had independent IQ scores, and, therefore, we were able to use IQ as an independent variable.

Results

Our analyses of the AFQT and ASVAB results showed the following.

- There was a strong association between AFQT scores and scores on each ASVAB component with level of education. (Table 1, Table 2).

- There appeared to be a differential effect of additional years of education in the direction of *lowered* test scores for African-Americans, women, and Hispanics and *higher* test scores for Whites and men. (Table 1).
- **Both patterns tended to persist even when IQ score was controlled.** (Table 2). That is, the highly significant partial regression coefficients of education (see Table 2 for Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and Mathematics) were not empirically consistent with the assumptions of the Herrnstein-Murray analysis.

Put another way, we found that the same patterns that Herrnstein and Murray reported for the “intellectual” components of the AFQT (Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and Mathematics) held for the other components when we used a model specification like theirs. However, when we considered specifications that tested whether there was a differential association between education level and AFQT test performance, we found that a model other than theirs provided a better explanation. Moreover, there was a highly significant contribution of education, even after controlling for IQ score.

We first ran regression analyses of the AFQT and its ASVAB components and obtained results very similar to those reported in Herrnstein and Murray. For example, this regression analysis of standard score form of the AFQT as reported, ZAFQTR, explained 44.4 percent of the variation in the standardized reported AFQT, a result significant at the 0.0001 level of significance.

The fitted equation was based on 3,712 respondents.

$$\begin{aligned}
 \text{ZAFQTR} = & -1.462 + 0.339(\text{ED_LVL80}) - 0.796(\text{BLACK}) - 0.470(\text{HISP}) \\
 & \qquad\qquad\qquad (674.93) \qquad\qquad\qquad (551.80) \qquad\qquad\qquad (144.96) \\
 & - 0.124(\text{AGE80}) + 0.0923(\text{INC80}) + 0.417(\text{MIL}) \\
 & \qquad\qquad\qquad (98.91) \qquad\qquad\qquad (108.69) \qquad\qquad\qquad (42.78) \\
 & - 0.0664(\text{SUPP}) - 0.146(\text{SEX}). \\
 & \qquad\qquad\qquad (4.71) \qquad\qquad\qquad (33.87)
 \end{aligned}$$

The value in parenthesis underneath each variable is the F-test of the null hypothesis that the variable’s partial regression coefficient is zero against the two-sided alternative that it is not zero. (An F-statistic for a partial regression coefficient with a value of 3.84 or greater is significant at the two-sided 0.05 level; a value

of 6.64 or greater is significant at the two-sided 0.01 level; and a value of 10.8 or greater is significant at the two-sided 0.001 level. We also ran parallel analyses using each ASVAB component as a dependent variable but do not report the complete results here).

The interpretation of this result is that respondents who were African-American scored 0.796 standard deviations below Whites, and Hispanic respondents scored 0.470 standard deviations lower than Whites. The F values of 551.80 for African-Americans and 144.96 are much higher than the value of 10.8 required for significance at the 0.001 level. Women scored 0.146 standard deviations lower than men, a highly significant difference with an F value of 33.87. (Herrnstein and Murray did not include sex in their specification). We found that the pattern of lower scores for African-Americans, Hispanics, and women held for the AFQT and most ASVAB components, consistent with Herrnstein and Murray (1994). We sought to identify additional factors that might clarify the reasons for the lower scores.

Differential Education Associations

We checked for underspecification of the first set of models by running a stepwise regression with a larger set of variables that consisted of all of the variables as well as all second and third order products of these variables. Since the interactions of income and education, age and education, sex and education, African-American and education, and Hispanic and education were significant for most of the ASVAB components, we considered a model with these variables added. The results are presented in Table 1. The first line contains the AFQT results. The next four lines contain the results for the components selected by Herrnstein and Murray as "intellectual". The bottom six lines contain the results for the remaining ASVAB components.

The partial regression coefficient of the education level that the subject reported in 1980 was again the most significant coefficient for the AFQT and each ASVAB component. Family income also had significant positive associations with the AFQT and each ASVAB component. The interaction of education level and income had a significant negative association with the AFQT and each ASVAB component. The interactions of African-American

Table 1
 Multiple Regression Results with Interactions for AFQT and ASVAB Components Controlling for Education, Age, Income, Sample Selection, Sex, and Ethnicity

Variable	R	ED_LVL80	AGE80	INC80	MIL	SUPP	SEX	BLACK	HISP	INC_ED	SEX_ED	BLACK_ED	HISP_ED
ZAFQTR	0.456 (257.97)	0.423*** (700.55)	-0.115*** (86.24)	0.416*** (65.98)	0.393*** (38.64)	-0.066* (4.75)	0.498*** (12.62)	-0.039 (0.05)	0.138 (0.45)	-0.027*** (40.69)	-0.056*** (21.82)	-0.065*** (17.79)	-0.053*** (8.91)
Total AFQT	0.396 (202.35)	0.091*** (456.00)	-0.024*** (54.00)	0.090*** (43.87)	0.093*** (30.56)	-0.014 (2.86)	0.029 (0.60)	0.003 (0.00)	0.054 (0.97)	-0.006*** (28.63)	-0.009** (7.60)	-0.019*** (20.76)	-0.016** (10.59)
Reasoning	0.421 (224.29)	0.094*** (494.18)	-0.026*** (60.85)	0.132*** (93.48)	0.137*** (66.21)	-0.018* (4.94)	0.042 (1.28)	-0.254*** (28.38)	-0.127* (5.45)	-0.009*** (64.94)	-0.005 (2.20)	0.005 (1.46)	0.002 (0.11)
World Knowledge	0.342 (160.06)	0.098*** (451.33)	-0.032*** (79.37)	0.126*** (73.04)	0.129*** (50.21)	-0.017* (3.84)	0.166*** (16.85)	-0.192*** (13.69)	-0.141* (5.70)	-0.009*** (50.36)	-0.012*** (12.19)	0.002 (0.25)	0.004 (0.52)
Paragraph Comp.	0.343 (161.25)	0.105*** (551.30)	-0.033*** (89.68)	0.107*** (55.59)	0.068*** (14.52)	-0.014 (2.60)	0.142*** (12.94)	0.104* (4.25)	0.049 (0.71)	-0.007*** (34.13)	-0.015*** (20.29)	-0.023*** (27.89)	-0.013** (6.64)
Math	0.402 (207.04)	0.092*** (438.81)	-0.024*** (50.93)	0.102*** (52.69)	0.098*** (31.87)	-0.016 (3.64)	0.065 (2.87)	-0.156** (10.03)	-0.126* (5.00)	-0.007*** (35.88)	-0.014*** (19.08)	-0.004 (0.98)	-0.001 (0.02)
General Science	0.281 (120.54)	0.089*** (374.87)	-0.032*** (78.41)	0.101*** (47.29)	0.150*** (68.10)	-0.002 (0.07)	0.202*** (25.03)	-0.142** (7.59)	-0.100 (2.86)	-0.007*** (32.59)	-0.013*** (15.12)	-0.000 (0.01)	0.002 (0.21)
Numerical Operations	0.298 (130.99)	0.082*** (314.55)	-0.021*** (35.36)	0.082*** (30.63)	0.132*** (52.57)	-0.012 (1.98)	0.217*** (28.87)	-0.026 (0.25)	-0.009 (0.02)	-0.006*** (21.21)	-0.011** (10.24)	-0.011* (6.16)	-0.004 (0.67)
Coding Speed	0.487 (292.88)	0.052*** (192.84)	-0.005 (3.03)	0.063*** (27.33)	0.089*** (35.82)	-0.012 (2.77)	-0.037 (1.23)	-0.195*** (21.35)	-0.115** (5.66)	-0.004** (20.18)	-0.017*** (35.04)	-0.001 (0.15)	0.000 (0.01)
Auto and Shop	0.414 (218.13)	0.070*** (285.66)	-0.013*** (16.40)	0.081*** (36.79)	0.077*** (22.26)	-0.017* (4.70)	0.010 (0.08)	-0.056 (1.44)	0.000 (0.00)	-0.006*** (27.44)	-0.016*** (26.08)	-0.014*** (12.85)	0.010* (4.42)
Mechanical Comp.	0.430 (232.99)	0.074*** (305.89)	-0.011*** (11.62)	0.079*** (33.91)	0.105*** (38.76)	-0.016* (3.95)	0.028 (0.56)	-0.037 (0.59)	-0.116* (4.49)	0.005*** (23.68)	-0.018*** (31.01)	-0.015*** (13.07)	-0.003 (0.34)
Electronics													

N=3712 Oldest sibling only, if more than one taking survey. (F-test value in parentheses.)

*p<.05** p<.01*** p<.001, F-test.

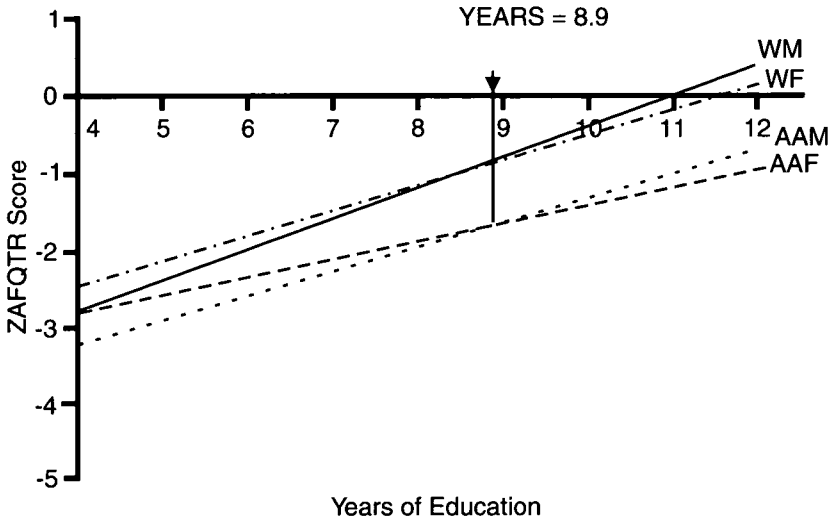
and education level (BLACK_ED) and sex and education level (SEX_ED) were also significantly negative in the Total AFQT score. A plot best shows the implications of these associations.

Figure 1 contains a plot of the fitted value of the standardized AFQT score for African-American subjects, both male and female, compared to White subjects, both male and female for each level of education when family income was set to \$15,000, age in 1980 was set to 20, and SUPP and MIL indicator variables were set to 0. Controlling for sex, African-American subjects at all levels of education did not score as high as comparable White subjects, as shown by the fact that the African-American line for a given sex is always below the line for Whites of the same sex. Controlling for race, female subjects with about nine or more years of education did not score as highly on the AFQT as equally well educated male subjects, as shown by the intersection of the male line of an ethnicity with the female line at around nine years of education. Specifically, the years of education that produce equal fits for women and men is calculated by dividing the SEX regression coefficient (0.498) by the negative of the SEX_ED regression coefficient (0.056) to obtain an education level of 8.9 years.

An examination of the results in Table 1 shows that the pattern followed by African-American males in the ZAFQTR held for five of the ASVAB components. There were significant results for African-Americans for each test in the sense that one or both of the partial regression coefficients for BLACK and BLACK_ED were significant at the 0.05 level or smaller. For example, the least significant finding was that the Coding Speed regression coefficient for BLACK_ED had an F statistic value of 6.16, significant at the 0.05 level. For African-Americans, there were highly significant negative differential associations with education (see column 12 titled BLACK_ED) for Arithmetic Reasoning (F=20.76), Mathematics (F=27.89), Coding Speed (F=6.16), Mechanical Comprehension (F=12.85), and Electronics (F=13.07). In summary, for all education levels and all tests, the fitted test score value for an African-American male was less than the fitted test score for a comparable White male.

The negative differential associations with Hispanics were much less significant (see column 13 titled HISP_ED) but roughly parallel the findings for African-Americans. There were signifi-

Figure 1
 Comparison Between ZAFQTR Score and Years of Education (White vs Black)



cant differences in test score improvement per year of education against Hispanics in the Total AFQT, Arithmetic Reasoning, and Mathematics components.

There were highly significant negative differential associations with education for women (see column 11 titled SEX_ED) for all ASVAB components except for Word Knowledge. As in the total AFQT results shown in Figure 1, while the coefficients of SEX_ED documented less improvement in test scores per year of additional education for White women compared to White men, the fitted test score for White women would be higher than the fitted test score for a White male at lower levels of education. Consequently, we calculated the education level at which a White woman would have the same fitted test score as a comparable White male. White women had higher fitted test scores than comparable White men on Paragraph Comprehension for education level up to 14 years, on Numerical Operations for education level up to 16 years, and on Coding Speed for education level up to 20 years. That is, White women up to the level of an associate degree had higher tests scores on these three components.

On the remaining components, however, White women did not do as well as comparable White men with higher levels of education. For example, a White woman had lower fitted tests scores than a comparable White male for Arithmetic Reasoning for an education level of 3 years or more, on Mathematics for an education level of 10 years or more, on General Science for an education level of 5 years or more, on Auto and Shop for an education level of 2 years or more, on Mechanical Comprehension for an education level of 1 year or more, and on Electronics for an education level of 2 years or more.

Controlling for IQ Score

We also ran regression analyses using the AFQT and ASVAB components as dependent variables and added the subject's IQ score and the interactions of education with sex and minority status for the 666 subjects for whom we had independent IQ scores. We report the results in Table 2. As before, the first line contains results for the total AFQT. The next four lines contain the results for the components used in the Herrnstein-Murray analysis. The partial regression coefficients in the fourth column (titled ED.LVL80, education level controlling for the other independent variables; that is, including IQ) should have been zero if the Herrnstein-Murray assumptions were empirically valid. As discussed below, these coefficients were highly significant. The next six lines contain the results for the other ASVAB components. Because of the small number of cases, we do not use the indicator variables MIL and SUPP. We also do not use the Income-education interaction variable. We use the indicator variable MINORITY, which is (1) if the subject is BLACK or HISPANIC and the product of MINORITY and Education.

For brevity, we list the ZAFQTR results. The fraction of the variance explained in the ZAFQTR score was 0.71, and the fitted regression model for the ZAFQTR was

$$\begin{aligned}
 & -1.973 + 0.623(\text{Z_MC_IQ}) + 0.148(\text{ED_LVL_80}) + 0.024(\text{AGE80}) \\
 & \quad (650.27) \quad (35.47) \quad (1.24) \\
 & + 0.034(\text{INC80}) + 0.247(\text{SEX}) + 0.131(\text{MINORITY}) \\
 & \quad (5.16) \quad (1.04) \quad (0.21) \\
 & - 0.035(\text{MIN_ED_LVL_80}) - 0.036(\text{SEX_ED_LVL_80}) \\
 & \quad (2.14) \quad (3.22)
 \end{aligned}$$

Table 2
 Multiple Regression Results on AFQT and ASVAB Components Controlling for IQ, Sex, Education, Income, Age, and Ethnicity

Dependent Variable	R ²	Z_MC_IQ	ED_LVL80	AGE80	INC80	SEX	MINORITY	MIN_ED	SEX_ED
ZAFQTR	0.71	0.623*** (650.27)	0.148*** (35.47)	0.024 (1.24)	0.034* (5.16)	0.247 (1.04)	0.131 (0.21)	-0.035+ (2.14)	-0.036+ (3.22)
Total AFQT									
ZASVAB2	0.61	0.149*** (408.72)	0.025*** (11.55)	0.007 (1.09)	0.043 (0.90)	-0.110 (1.09)	0.012 (0.02)	-0.008 (1.29)	0.003 (0.22)
Arithmetic Reasoning									
ZASVAB3	0.59	0.120*** (299.03)	0.026*** (13.44)	0.011+ (2.93)	0.082+ (3.18)	0.015 (0.05)	-0.212** (6.92)	0.009+ (1.94)	-0.002 (0.14)
Word Knowledge									
ZASVAB4	0.49	0.112*** (200.89)	0.034*** (18.39)	0.006 (0.76)	0.013** (7.04)	0.202** (6.66)	-0.078 (0.72)	0.001 (0.01)	-0.143* (4.93)
Paragraph Comp.									
ZASVAB8	0.57	0.154*** (397.15)	0.034*** (18.90)	-0.000 (0.00)	0.118* (6.27)	0.068 (0.77)	0.066 (0.54)	-0.009 (1.50)	-0.009+ (1.90)
Math									
ZASVAB1	0.56	0.126*** (272.94)	0.026*** (11.46)	0.007 (0.94)	0.071+ (2.31)	-0.026 (0.12)	-0.197* (4.94)	0.009 (1.32)	-0.008+ (1.85)
General Science									
ZASVAB5	0.35	0.097*** (130.98)	0.027** (9.62)	0.002 (0.09)	0.081+ (2.45)	0.166* (3.94)	-0.067 (0.46)	0.002 (0.08)	-0.009+ (1.87)
Numerical Operations									
ZASVAB6	0.38	0.079*** (90.60)	0.037*** (19.16)	0.003 (0.19)	0.078+ (2.35)	0.255** (9.54)	-0.010 (0.01)	-0.006 (0.52)	-0.015* (4.74)
Coding Speed									
ZASVAB7	0.53	0.061*** (62.92)	0.014+ (3.21)	0.012+ (2.69)	0.067+ (2.06)	-0.104 (1.88)	-0.316*** (12.64)	0.014+ (3.68)	-0.013* (4.05)
Auto and Shop									
ZASVAB9	0.50	0.099*** (142.47)	0.026** (9.28)	0.006 (0.76)	0.064 (1.63)	-0.069 (0.71)	-0.059 (0.38)	-0.005 (0.36)	-0.011+ (2.59)
Mech. Comp.									
ZASVAB10	0.52	0.097*** (141.37)	0.024** (8.09)	0.010+ (2.02)	0.015 (0.10)	-0.040 (0.24)	-0.150+ (2.50)	0.001 (0.03)	-0.014* (4.28)
Electronics									

N=666 Oldest sibling only, if more than one taking survey.
 +p<.10 one-sided test, *p<.05 two-sided, ** p<.01 two-sided, *** p<.001, two-sided test.

As expected, the partial regression coefficient of the subject's reported IQ score had the greatest F statistics in each regression. Contrary to expectation, the subject's education level also had a significant partial regression coefficient in each regression, except for Auto and Shop. The significant association of educational level with the ASVAB components used in the Herrnstein and Murray analysis is inconsistent with the claim that these measures are "IQ scores." Additionally, the significant associations of income with both Paragraph Comprehension ($F=7.04$) and Mathematics ($F=6.27$) after controlling for IQ are also not consistent with the Herrnstein-Murray assumptions.

If the differential education models that we fit in the previous section were valid, then the magnitude of the partial regression coefficients of the interactions should remain roughly the same after the subject's reported IQ score is included. The indications of differential test score improvement in the Total AFQT against women ($p<0.05$, one-sided test) remained even though reported IQ was entered and the number of cases used in the analysis was greatly reduced. Women had fitted AFQT scores less than comparable men when the education level was 7 years or more (compared to 9 years above). Women also had lower fitted test scores than comparable men (significant with 0.05 two-sided test) in the Paragraph Comprehension component for education level greater than 1 year, Auto and Shop for all education levels, and Electronics for all education levels.

There was still a trend ($p<0.10$, one sided test) for minority subjects to have less test score improvement per year of education in the Total AFQT, even with IQ controlled. Minority subjects with education level 4 years or more did more poorly than comparable White males. After controlling for IQ, minority test scores were lower than White test scores in the Word Knowledge component for education level less than 24 years, in the General Science component for all education levels, and in the Auto and Shop component at all education levels.

Conclusions and Discussion

This study focuses on the relationship between years of education, race, gender, and the acquisition of marketable skills. Using

data from a nationally representative sample of adolescents and young adults, the analyses concentrated on two questions. First, do African-Americans and women score lower on a standardized test of verbal, mathematical, and technical skills? The answer is yes they do. Second, do the African-American-White and male-female gaps in test scores remain about the same for different levels of educational attainment? The answer is no; there is a general pattern that women have less test score improvement with an additional year of education. This pattern of association holds even with IQ controlled. A similar pattern holds for African-Americans. More specifically, there appears to be less improvement in test scores for each year of increased education level in women and minority men than for males who reported themselves as other than African-American or Hispanic. These disparities are most apparent in quantitative and technical areas. The NLSY data show that, on a societal-wide basis, there was unequal test score improvement per year of education prior to 1980 that accounts for much of the difference between minority and White performance on the AFQT.

Our findings support and expand upon results reported by other researchers (e.g., Fischer *et al.* 1996; Korenman and Winship 1995; Phillips, Crouse, and Ralph 1998). All investigations document the higher scores of men relative to women and Whites compared to African-Americans on standardized tests of academic achievement. Our analysis expands these findings by specifying a regression model that explicitly tests whether (or not) the AFQT score gap widens with increasing years of education and documents that the pattern remains after IQ is controlled.

The significance of the partial regression coefficient of the interaction of education level and being African-American or Hispanic is, *ipso facto*, a significant issue to be considered in a complete understanding of the differences in subjects' performances on the AFQT and its components. The findings in *The Bell Curve* of decreased economic potential for subjects whose AFQT scores were lower are documentation of the costs of differential test score improvement. It is not surprising that Korenman and Winship (1995) have confirmed this portion of findings in *The Bell Curve*; namely, that, subjects who do not test well with respect to

having the qualifications needed in a technological society are at a disadvantage economically.

Two limitations should be noted with regard to the present study. While the results show that improvements in test scores per year of education for women and minorities are less than for males and Whites with each additional year of schooling, we cannot determine the source of the disparity. Past research suggests that, disadvantaged background, poverty, parenting style, neighborhood (Korenman and Winship 1995; Phillips *et al.* 1998), low achieving peers (Jencks and Phillips 1998b), and a poor school environment, including teacher expectations, poor funding, and larger classes (Ferguson 1998a; 1998b) affect the acquisition of marketable skills. Future research needs to identify exactly which factors best explain the widening gap that emerges in test scores with increasing years of education. The second limitation is with the sample of 666 for which there are independent IQ scores. One cannot exclude the existence of an unknown selection bias whose effects are unpredictable.

Nevertheless, in terms of policy implications, our results show that there is a differential improvement in test scores with each additional year of education that works against women (specifically in the AFQT, Mathematics, General Science, Auto and Shop, Mechanical Comprehension, and Electronic components) and against African-Americans (specifically in the AFQT, Arithmetic Reasoning, Mathematics, Coding Speed, Mechanical Comprehension, and Electronic components). Further, this differential lack of improvement appears to explain the reduced scores for African-Americans reported by Herrnstein and Murray.

Consequently, we argue that there exists much greater opportunity for positive intervention to reduce the observed test score differences between Whites and others than Herrnstein and Murray (1994) lead their readers to believe possible. Our results suggest that these efforts should focus on the various dimensions of the educational process.

References

- Arons, S. (1997). *Short Route to Chaos: Conscience, Community, and the Re-Constitution of American Schooling*. Amherst, MA: The University of Massachusetts Press.

- Bane, M. J., & D. B. Ellwood, (1994). *Welfare Reform*. New York: Russell Sage Foundation.
- Center for Human Resource Research (1994). *NLS Handbook, 1994, The National Longitudinal Surveys*. Columbus, OH: The Ohio State University.
- Coleman, J. S., et al., (1966). *Equality of Educational Opportunity*. Washington, D.C.: Government Printing Office.
- Coleman, J. S. (1989). "Response to Sociology of Education Section Award." *Footnotes: Official Newsletter of the American Sociological Association*, 17: 4–5.
- Currie, J., & Thomas, D. (1995). *Race, Children's Cognitive Achievement and The Bell Curve*. Cambridge, MA: National Bureau of Economic Research, Working Paper 5240.
- Cutright, P. (1974). "Academic Achievement, Schooling, and the Earnings of White and Black Men." *Journal of Social and Behavioral Sciences*, 20, 1–18.
- Devlin, B., Fienberg, S. E., Resnick, D. P., & Roeder, K. (1995a). "Galton Redux: Eugenics, Intelligence, Race, and Society: A Review of The Bell Curve: Intelligence and Class Structure in American Life." *Journal of the American Statistical Association*, 90, 1483–1488.
- Devlin, Bernie, Fienberg, Stephen E., Resnick, Daniel P., and Roeder, Kathryn. (1995b). "Wringing The Bell Curve: A Cautionary Tale About the Relationships Among Race, Genes, and IQ." *Chance*, 8, 27–36.
- Falconer, D.S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, Fourth edition. Essex, England: Longman.
- Ferguson, Ronald F. (1998a) "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." Pp. 273–145 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institutions Press.
- Ferguson, Ronald F. (1998b) "Can Schools Narrow the Black-White Test Score "Gap?" Pp. 318–374 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institutions Press.
- Fischer, Claude S., Hout, Michael, Jankowski, Martin Sanchez, Lucas, Samuel R., Swidler, Ann, and Voss, Kim. (1996). *Inequality by Design: Cracking the Bell Curve Myth*. Princeton, NJ: Princeton University Press.
- Goldberger, Arthur, Manski, Charles. (1995). "Review Article: "The Bell Curve by Herrnstein and Murray." *The Journal of Economic Literature*, 33, 762–776.
- Gould, Stephen J. (1994). "Curveball." *The New Yorker*. (November 28): 139–149.
- Grassmmer, David, Flanagan, Ann, and Williamson, Stephanie. (1998). "Why Did the Black-White Test Gap Narrow in the 1970's and 1980's?" Pp.182–226 in Christopher Jencks and Meredith Phillips (Eds.). *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institute Press.
- Hauser, Robert M. (1995). "Review of *The Bell Curve*." *Contemporary Sociology*, 24, 149–53.
- Heckman, James J. (1995). Lessons from *The Bell Curve*." *Journal of Political Economy*, 103, 1091–1122.
- Herrnstein, Richard J., Murray, Charles. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: The Free Press.
- Jencks, Christopher, and Phillips, Meredith. (1998a). "American's Next Achieve-

- ment Test: Closing the Black-White Test Score Gap." *The American Prospect*, No. 44 (September-October), 44-53.
- Jencks, Christopher, and Phillips, Meredith. (1998b). "The Black-White Test Score Gap: An Introduction." Pp. 1-51 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institutions Press.
- Jensen, Arthur, R. (1969). "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review*, 39, 1-123.
- Jensen, Arthur, R. (1985). "The Nature of Black-White Differences on Various Psychometric Tests: Spearman's Hypothesis." *Behavioral and Brain Sciences*, 8, 193-258.
- Kane, Thomas. (1998). "Racial and Ethnic Preferences in College Admission. Pp. 431-456 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institute Press.
- Katz, Michael. (1997). *In the Shadow of the Poor House*. (10th Anniversary Edition), New York: Basic Books.
- Kohn, Melvin L. (1996). "The 'Bell Curve' from the Perspective of Research on Social Structure and Personality." *Sociological Forum*, 11, 395-411.
- Korenman, Sanders, Winship, Christopher. (1995). "A Reanalysis of *The Bell Curve*." Cambridge, MA: National Bureau of Economic Research, Working Paper 5230.
- Massey, Douglas S., and Denton, Nancy. (1993). *American Apartheid: Segregation and the Making of the American Underclass*. Cambridge, MA: Harvard University Press.
- Mead, Lawrence. (1993). *The New Politics of Poverty: The Non-Working Poor in America*. New York: Basic Books.
- Mutaner, C. Nieto, F. J., O'Campo, P. (1996). "The 'Bell Curve', on Race, Social Class, and Epidemiological Research." *American Journal of Epidemiology*, 144, 531-536.
- Phillips, Meredith, Brooks-Gunn, Jeanne, Duncan, Greg J., Klebanov, Pamela, and Crane, Jonathan. (1998). "Family Background, Parenting Practices, and the Black-White Test Score Gap." Pp. 103-145 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institutions Press.
- Phillips, Meredith, Crouse, James, and Ralph, John. (1998). "Does the Black-White Test Score Gap Widen After Children Enter School?" Pp. 229-272 in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institutions Press.
- Rosenthal, Robert, and Jacobson, Lenore. (1968). *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart, and Wilson.
- Wilson, William Julius. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: The University of Chicago Press.
- Wilson, William Julius. (1996). *When Work Disappears: The World of the Working Poor*. New York: Knopf Publishers.