

CHRISTOPHER JENCKS  
MEREDITH PHILLIPS

# 1

## *The Black-White Test Score Gap: An Introduction*

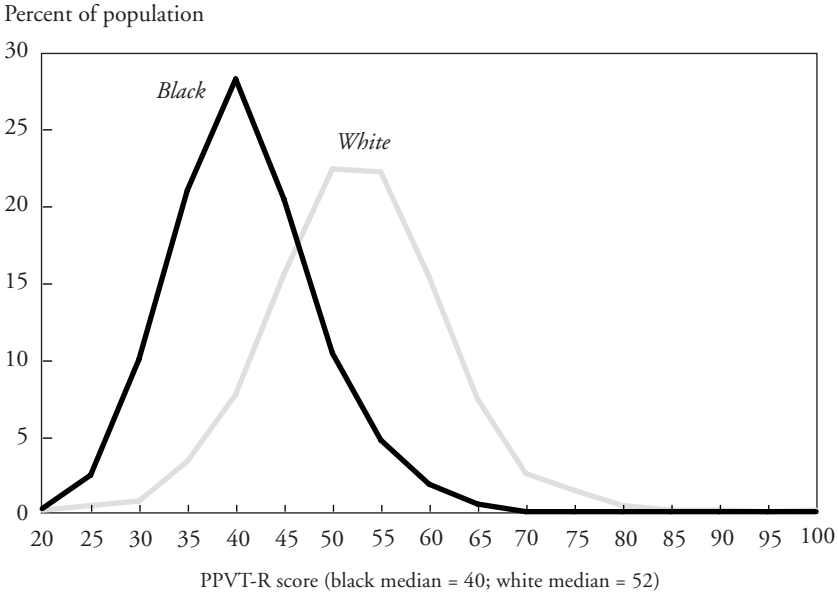
**A**FRICAN AMERICANS currently score lower than European Americans on vocabulary, reading, and mathematics tests, as well as on tests that claim to measure scholastic aptitude and intelligence.<sup>1</sup> This gap appears before children enter kindergarten (figure 1-1), and it persists into adulthood. It has narrowed since 1970, but the typical American black still scores below 75 percent of American whites on most standardized tests.<sup>2</sup> On some tests the typical American black scores below more than 85 percent of whites.<sup>3</sup>

1. We are indebted to Karl Alexander, William Dickens, Ronald Ferguson, James Flynn, Frank Furstenberg, Arthur Goldberger, Tom Kane, David Levine, Jens Ludwig, Richard Nisbett, Jane Mansbridge, Susan Mayer, Claude Steele, and Karolyn Tyson for helpful criticisms of earlier drafts. But we did not make all the changes they suggested, and they are in no way responsible for our conclusions.

2. These statistics also imply, of course, that a lot of blacks score above a lot of whites. If the black and white distributions are normal and have the same standard deviation, and if the black-white gap is one (black or white) standard deviation, then when we compare a randomly selected black to a randomly selected white, the black will score higher than the white about 24 percent of the time. If the black-white gap is 0.75 rather than 1.00 standard deviations, a randomly selected black will score higher than a randomly selected white about 30 percent of the time.

3. Although this book concentrates on the black-white gap, similar issues arise when we compare either Hispanics or Native Americans to whites or Asians. We have concentrated on the black-white gap because far more is known about test performance among blacks and whites than among other groups. This reflects the fact that white American scholars have traditionally been more concerned

Figure 1-1. *Vocabulary Scores for Black and White Three- and Four-Year-Olds, 1986–94*



Source: National Longitudinal Survey of Youth Child Data, 1986–94. Black N = 1,134; white N = 2,071. Figure is based on black and white three- and four-year-olds in the Children of the National Longitudinal Survey of Youth (CNLSY) data set who took the Peabody Picture Vocabulary Test-Revised (PPVT-R). The test is the standardized residual, coded to a mean of 50 and a standard deviation of 10, from a weighted regression of children's raw scores on their age in months, age in months squared, and year-of-testing dummies. See chapter 4 for details on the CNLSY and the PPVT-R.

The black-white test score gap does not appear to be an inevitable fact of nature. It is true that the gap shrinks only a little when black and white children attend the same schools. It is also true that the gap shrinks only a little when black and white families have the same amount of schooling, the same income, and the same wealth. But despite endless speculation, no one has found genetic evidence indicating that blacks have less innate intellectual ability than whites. Thus while it is clear that eliminating the test score gap would require enormous effort by both blacks and whites and would probably take more than one generation, we believe it can be done.

---

about the plight of blacks than about the plight of other minorities, as well as the fact that blacks were (until recently) far more numerous than Native Americans, Asians, or even Hispanics.

This conviction rests mainly on three facts:

—*When black or mixed-race children are raised in white rather than black homes, their preadolescent test scores rise dramatically.* Black adoptees' scores seem to fall in adolescence, but this is what we would expect if, as seems likely, their social and cultural environment comes to resemble that of other black adolescents and becomes less like that of the average white adolescent.<sup>4</sup>

—*Even nonverbal IQ scores are sensitive to environmental change.* Scores on nonverbal IQ tests have risen dramatically throughout the world since the 1930s.<sup>5</sup> The average white scored higher on the Stanford-Binet in 1978 than 82 percent of whites who took the test in 1932.<sup>6</sup> Such findings reinforce the implications of adoption studies: large environmental changes can have a large impact on test performance.

—*Black-white differences in academic achievement have also narrowed throughout the twentieth century.* The best trend data come from the National Assessment of Educational Progress (NAEP), which has been testing seventeen-year-olds since 1971 and has repeated many of the same items year after year. Figure 1-2 shows that the black-white reading gap narrowed from 1.25 standard deviations in 1971 to 0.69 standard deviations in 1996. The math gap fell from 1.33 to 0.89 standard deviations.<sup>7</sup> When Min-Hsiung Huang and Robert Hauser analyzed vocabulary scores for adults born between 1909 and 1969, the black-white gap also narrowed by half.

In a country as racially polarized as the United States, no single change taken in isolation could possibly eliminate the entire legacy of slavery and Jim Crow or usher in an era of full racial equality. But if racial equality is America's goal, reducing the black-white test score gap would probably do

4. See chapter 3.

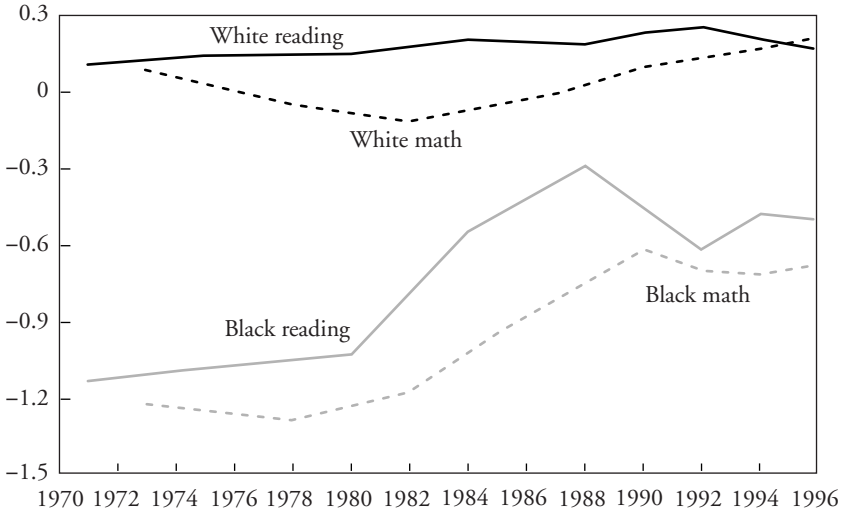
5. Flynn (1987); Neisser (1998).

6. The 1932–78 comparison for whites is derived from Flynn (1984), who shows that IQ scores rose by roughly 13.8 points between 1931–33 and 1976–80. Tuddenham (1948) reported an equally dramatic improvement in the test performance of American soldiers between World War I and World War II. Since the trend found by Flynn appears to have been roughly linear, the net increase between 1917 and 1978 was probably close to 18 points (1.2 standard deviations). Flynn (personal communication) reports that restandardization of the Wechsler intelligence tests shows a further increase in mean IQ since 1978.

7. The standardized black-white gap always looks larger in NAEP than in other surveys, because NAEP reports standard deviations that have been corrected for measurement error. In addition, the standard deviation of seventeen-year-olds' reading and math scores has fallen over time, making the pre-1996 gaps look larger than they would if they were expressed in contemporary standard deviations. Five other major national surveys of high school seniors conducted since 1965 also show black-white convergence (see chapter 5), as do surveys of younger children (see chapter 6 and chapter 7).

Figure 1-2. *NAEP Reading and Mathematics Scores for Black and White Seventeen-Year-Olds, 1971–96*

Standardized score using 1996 mean and SD



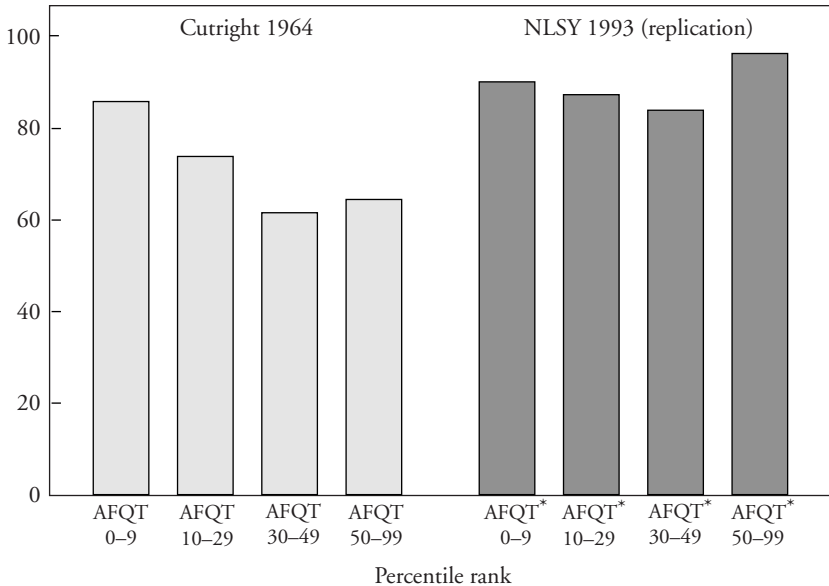
Source: National Assessment of Educational Progress. Tests in all years are in a common metric and have been rescaled so that the 1996 population mean is zero and the 1996 standard deviation is 1.00.

more to promote this goal than any other strategy that commands broad political support. Reducing the test score gap is probably both necessary and sufficient for substantially reducing racial inequality in educational attainment and earnings. Changes in education and earnings would in turn help reduce racial differences in crime, health, and family structure, although we do not know how large these effects would be.

This judgment contradicts Christopher Jencks and his colleagues' 1972 conclusion in *Inequality* that reducing cognitive inequality would not do much to reduce economic inequality. The reason is simple: the world has changed. In 1972 the best evidence about what happened to black workers with high test scores came from a study by Phillips Cutright, who had analyzed the 1964 earnings of men in their thirties who had taken the Armed Forces Qualification Test (AFQT) between 1949 and 1953.<sup>8</sup> Over-

8. Cutright (1972, 1974). Cutright's sample, which was 30 to 37 years old in 1964, included both veterans and nonveterans. His checks suggested that it was quite representative of men in this age

Figure 1-3. *Black Annual Earnings as a Percentage of White Earnings among Employed Men Aged 30 to 37 in 1964 or 31 to 36 in 1993, by Percentile Rank on a Military Test Taken When the Men Were Aged 18 to 23* Percent



Sources: Cutright (1972) and authors' tabulations from the NLSY. Cutright's version of the AFQT included tests for vocabulary, arithmetic, and spatial relations. Our NLSY approximation of his AFQT included tests for word knowledge, numerical operations, and mechanical reasoning (AFQT\*). See the notes in the text for details on the samples and standard errors.

all, employed black men earned 57.5 percent of what whites earned. Among men with AFQT scores above the national average, black men earned 64.5 percent of what whites earned (figure 1-3).<sup>9</sup> In such a world, eliminating racial differences in test performance did not seem likely to reduce the earnings gap very much.

---

group. His earnings data came from the Social Security Administration and were therefore limited to those in covered employment. Social security withholding covered only the first \$4,800 of earnings in 1964, so Cutright estimated the total 1964 earnings of those who reached the \$4,800 ceiling by assuming that they continued to earn the amount they received in the quarter before they reached the ceiling.

9. Using five AFQT intervals, Cutright (1972) reported mean annual earnings for black and white men who worked. We collapsed the top two intervals, because Cutright's sample of 4,051 employed blacks included only 53 men in his top AFQT interval (80th to 99th percentile), and because our

Today's world is different. The best recent data on test scores and earnings come from the National Longitudinal Survey of Youth (NLSY), which gave the Armed Services Vocational Aptitude Battery to a national sample of young people in 1980.<sup>10</sup> Among employed men who were 31 to 36 years old in 1993, blacks earned 67.5 percent of what whites earned, a modest but significant improvement over the situation in 1964.<sup>11</sup> The big change occurred among blacks with test scores near or above the white average. Among men who scored between the 30th and 49th percentiles nationally, black earnings rose from 62 to 84 percent of the white average. Among men who scored above the 50th percentile, black earnings rose from 65 to 96 percent of the white average.<sup>12</sup> In this new world, raising black workers' test scores looks far more important than it did in the 1960s.<sup>13</sup>

comparison sample of 736 black men from the National Longitudinal Survey of Youth includes only 21 men in this interval.

10. Almost all members of Cutright's sample took a version of the AFQT that included equal numbers of questions on vocabulary, arithmetic, and spatial relations. The Armed Services Vocational Aptitude Battery (ASVAB) did not include a spatial relations test, so we could not exactly reproduce the 1950–53 AFQT. To approximate the 1950–53 test we summed men's scores on the ASVAB tests of word knowledge, numerical operations, and mechanical reasoning. Our results did not change appreciably when we used other ASVAB tests instead.

11. To match Cutright's sample, we selected men who took the ASVAB between the ages of 18 and 23. Cutright had no data on whether respondents were Hispanic. We treated Hispanics as white, but our results hardly changed when we dropped Hispanics. Cutright's data were limited to men who worked in jobs covered by social security in 1964, whereas our NLSY sample covered all men of the relevant age.

12. Cutright (1972) reported the mean earnings of black and white men with AFQT scores in each interval. He also reported the linear regression of men's unlogged earnings on their AFQT percentile score (coded 0 to 99). We estimated a similar regression. To express coefficients in a common metric, we divided each coefficient by the mean earnings of the relevant group in the relevant year. The regression coefficients (B), standard errors (SE), and unweighted sample sizes (N) were as follows:

	Blacks <u>1964</u>	Blacks <u>1993</u>	Whites <u>1964</u>	Whites <u>1993</u>
B	.52	1.33	.69	.82
(SE)	(.11)	(.11)	(.03)	(.05)
N	4,051	7,362	5,022	1,842

13. ASVAB scores explain more of the racial gap in annual earnings in figure 1-3 than in chapter 14 by William Johnson and Derek Neal. This discrepancy has at least six sources, most of which derive from our effort to maintain comparability with Cutright. First, unlike Johnson and Neal we include respondents with military earnings. Second, whereas Johnson and Neal include anyone with positive earnings between 1990 and 1992, we include only individuals with positive earnings in 1993, which excludes more black men with weak labor force attachment. Third, unlike Johnson and Neal, we top-coded earnings (at 3.16 times the white mean, as in Cutright), raising the black-white earnings ratio by about 0.03. Fourth, we used different ASVAB tests from Johnson and Neal, although this makes very little difference. Fifth, we included men from all the NLSY sampling strata and used sample weights, whereas Johnson and Neal excluded the poor white oversample and the military sample and

Some skeptics have argued that scores on tests of this kind are really just proxies for family background. As we shall see, family background does affect test performance. But even when biological siblings are raised in the same family, their test scores hardly ever correlate more than 0.5. Among children who have been adopted, the correlation falls to around half that level.<sup>14</sup> The claim that test scores are only a proxy for family background is therefore false. Furthermore, test score differences between siblings raised in the same family have sizable effects on their educational attainment and earnings.<sup>15</sup> Thus while it is true that eliminating the black-white test score gap would not reduce the black-white earnings gap quite as much as figure 1-3 implies, the effect would surely be substantial.

Reducing the black-white test score gap would reduce racial disparities in educational attainment as well as in earnings. The nationwide High School and Beyond survey tested twelfth-graders in 1982 and followed them up in 1992, when they were in their late twenties. At the time of the followup only 13.3 percent of the blacks had earned a B.A., compared with 30 percent of the non-Hispanic whites. Many observers blame this disparity on black parents' inability to pay college bills, black students' lack of motivation, or the hostility that black students encounter on predominantly white college campuses. All these factors probably play some role. Nonetheless, figure 1-4 shows that when we compare blacks and whites with the same twelfth grade test scores, blacks are *more* likely than whites to complete college. Once we equalize test scores, High School and Beyond blacks' 16.7 point disadvantage in college graduation rates turns into a 5.9 point advantage.<sup>16</sup>

Eliminating racial differences in test performance would also allow colleges, professional schools, and employers to phase out the racial preferences that have caused so much political trouble over the past generation. If selective colleges based their admission decisions solely on applicants' predicted college grades, their undergraduate enrollment would currently

---

did not weight their data. Sixth, we included men who took the ASVAB between the ages of 18 and 23, whereas Johnson and Neal included men who took it between the ages of 15 and 18, so our ASVAB scores are more likely than theirs to have been affected by the length of time men spent in school.

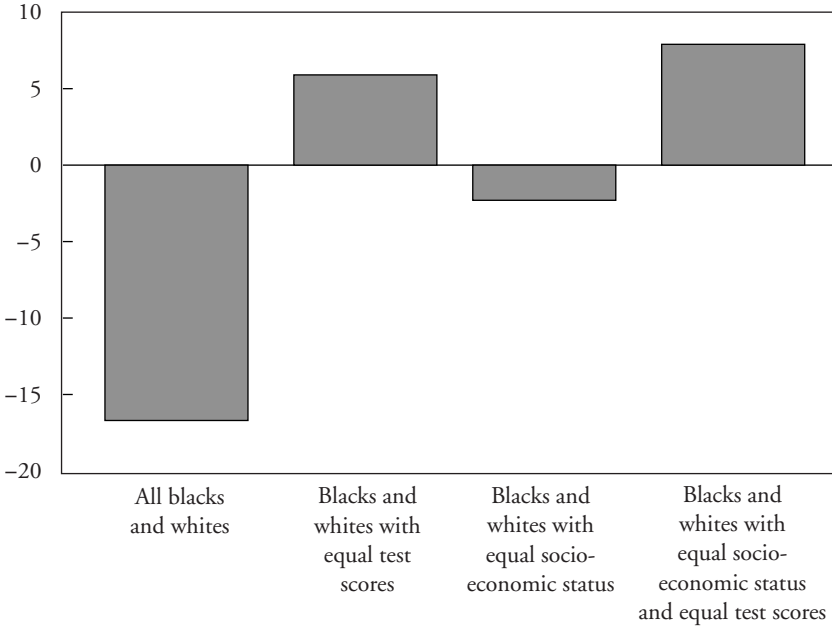
14. See Jencks and others (1972, appendix A); Loehlin (1980); and Cherny and Cardon (1994).

15. Korenman and Winship (forthcoming).

16. The results shown in figure 1-4 do not change if we use years of schooling rather than college graduation as the dependent variable. Similar results can be found in samples dating back to the early 1960s, so they are not attributable to affirmative action. Affirmative action mainly affects *where* blacks go to college, not *whether* they go.

Figure 1-4. *Gap in Eventual College Graduation Rates among Blacks and Whites Who Were in Twelfth Grade in 1982, Controlling Socioeconomic Status and Test Scores, 1992*

Black-white gap in percent with a B.A.



Source: Authors' tabulations from High School and Beyond 1992 followup. Test score is the sum of vocabulary, reading, and math scores. Socioeconomic status includes parents' income, occupation, schooling, possessions in the home, marital status, number of siblings, urbanism, and region. The standard error for black-white gap is about 2.5 percentage points.

be 96 or 97 percent white and Asian. To avoid this, almost all selective colleges and professional schools admit African Americans and Hispanics whom they would not admit if they were white. Racial preferences of this kind are politically unpopular.<sup>17</sup> If selective colleges could achieve racial diversity without making race an explicit factor in their admission decisions, blacks would do better in college and whites would nurse fewer political grudges.

17. Proposition 209 banned racial preferences in California's public institutions but did not cover private institutions. The Fifth Circuit's *Hopwood v. State of Texas* decision banned such preferences in both public and private institutions throughout Texas, Louisiana, and Mississippi. The U.S. Supreme Court refused to review *Hopwood*, so it does not apply elsewhere in the United States. Should the Supreme Court endorse the *Hopwood* principle in the future, it would cover both public and private institutions throughout the nation.



Advocates of racial equality might be more willing to accept our argument that narrowing the test score gap is crucial to achieving their goals if they believed that narrowing the gap was really feasible. But pessimism about this has become almost universal. In the 1960s, racial egalitarians routinely blamed the test score gap on the combined effects of black poverty, racial segregation, and inadequate funding for black schools. That analysis implied obvious solutions: raise black children's family income, desegregate their schools, and equalize spending on schools that remain racially segregated. All these steps still look useful, but none has made as much difference as optimists expected in the early 1960s.

—The number of affluent black parents has grown substantially since the 1960s, but their children's test scores still lag far behind those of white children from equally affluent families. Income inequality between blacks and whites appears to play some role in the test score gap, but it is quite small.<sup>18</sup>

—Most southern schools desegregated in the early 1970s, and southern black nine-year-olds' reading scores seem to have risen as a result.<sup>19</sup> Even today, black third-graders in predominantly white schools read better than initially similar blacks who have attended predominantly black schools. But large racial differences in reading skills persist even in desegregated schools, and a school's racial mix does not seem to have much effect on changes in reading scores after sixth grade or on math scores at any age.<sup>20</sup>

—Despite glaring economic inequalities between a few rich suburbs and nearby central cities, the average black child and the average white child now live in school districts that spend almost exactly the same amount per pupil.<sup>21</sup> Black and white schools also have the same average number of

18. See chapter 4.

19. See chapter 6.

20. Phillips (1997) found that a 10 point increase in the percentage of white students was associated with a .027 standard deviation increment in third-grade reading scores, holding "true" first-grade scores constant. Attending a 90 percent white school rather than an all-black school could therefore raise third-grade reading scores by .243 standard deviations. The estimated effect was smaller in later elementary grades and often changed signs. For math the effects were smaller than for reading and changed sign in high school.

21. Tabulations by William Evans show that the average black student lived in a district that spent \$5,387 in 1992, while the average white student lived in a district that spent \$5,397. In 1972 the figures (in 1992 dollars) were \$3,261 for blacks and \$3,397 for whites. These estimates come from a file that matches expenditure data collected by the Census of Governments in 1972, 1982, and 1992 with demographic data on Census tracts in 1970, 1980, and 1990. A few districts cannot be matched. For details on the samples see Evans, Murray, and Schwab (1997).

teachers per pupil, the same pay scales, and teachers with almost the same amount of formal education and teaching experience.<sup>22</sup> The most important resource difference between black and white schools seems to be that teachers in black schools have lower test scores than teachers in white schools. This is partly because black schools have more black teachers and partly because white teachers in black schools have unusually low scores.<sup>23</sup>

For all these reasons, the number of people who think they know how to eliminate racial differences in test performance has shrunk steadily since the mid-1960s. While many people still think the traditional liberal remedies would help, few now believe they would suffice.

Demoralization among liberals has given new legitimacy to conservative explanations for the test score gap. From an empirical viewpoint, however, the traditional conservative explanations are no more appealing than their liberal counterparts. These explanations fall into three overlapping categories: the culture of poverty, the scarcity of two-parent black families, and genes.

—In the 1960s and 1970s, many conservatives blamed blacks' problems on a culture of poverty that rejected school achievement, the work ethic, and the two-parent family in favor of instant gratification and episodic violence. In the 1980s, conservatives (as well as some liberals) characterized the "black underclass" in similar terms. But this description only fits a tiny fraction of the black population. It certainly cannot explain why children from affluent black families have much lower test scores than their white counterparts.

—Conservatives invoke the decline of the family to explain social problems almost as frequently as liberals invoke poverty. But once we control a mother's family background, test scores, and years of schooling, whether she is married has even less effect on her children's test scores than whether she is poor.<sup>24</sup>

22. On teacher-pupil ratios see chapter 9. On salaries, teacher training, and teacher experience, see Boozer, Krueger, and Wolkon (1992, p. 299). Boozer, Krueger, and Wolkon's data cover secondary schools. Because most teachers prefer high-scoring students, and experienced teachers get preference in assignments, we would expect blacks to get less experienced teachers than whites in the same high school. We have not seen comparable recent data on elementary schools. For earlier data see Coleman and others (1966) and Mosteller and Moynihan (1972).

23. See chapter 9 by Ronald Ferguson. Ferguson's data come from Alabama and Texas, but the same pattern almost certainly recurs in other states.

24. See chapter 4.

—Scientists have not yet identified most of the genes that affect test performance, so we have no direct genetic evidence regarding innate cognitive differences between blacks and whites. But we have accumulated a fair amount of indirect evidence since 1970. Most of it suggests that whether children live in a “black” or “white” environment has far more impact on their test performance than the number of Africans or Europeans in their family tree.<sup>25</sup>

Taken as a whole, then, what we have characterized as the “traditional” explanations for the black-white test score gap do not take us very far. This has led some people to dismiss the gap as unimportant, arguing that the tests are culturally biased and do not measure skills that matter in the real world. Few scholars who spend time looking at quantitative data accept either of these arguments, so they have had to look for new explanations of the gap. These new explanations can mostly be grouped under two overlapping headings: culture and schooling.

In the late 1960s and early 1970s, many blacks and some whites dismissed cultural explanations of the test score gap as an effort to put down blacks for not thinking and acting like upper-middle-class whites. Since then, cultural explanations have enjoyed a slow but steady revival. In 1978 the Nigerian anthropologist John Ogbu suggested that caste-like minorities throughout the world tended to do poorly in school, even when they were visually indistinguishable from the majority.<sup>26</sup> Later, Ogbu argued that because blacks had such limited opportunities in America, they developed an “oppositional” culture that equated academic success with “acting white.”<sup>27</sup> By linking black culture directly to oppression, Ogbu made it much easier for liberals to talk about cultural differences. Jeff Howard and Ray Hammond added another important strand to this argument when they suggested that academic competence developed partly through competition, and that “rumors of inferiority” made blacks reluctant to compete academically.<sup>28</sup> More recently, Claude Steele has argued that people of all races avoid situations in which they expect others to have negative stereotypes about them, even when they know that the stereotype does not apply. According to Steele, many black students “disidentify” with school because

25. See chapter 3.

26. Ogbu (1978).

27. Ogbu (1986); Fordham and Ogbu (1986). Philip Cook and Jens Ludwig discuss the empirical evidence for the “acting white” theory in chapter 10.

28. Howard and Hammond (1985).

constructing a personal identity based on academic competence entails a commitment to dealing with such stereotypes on a daily basis.<sup>29</sup>

Social scientists' thinking about "school effects" has also changed since the late 1960s. The 1966 Coleman Report and subsequent "production function" studies convinced most economists and quantitative sociologists that school resources had little impact on achievement.<sup>30</sup> Since 1990, however, new statistical methods, new data, and a handful of genuine experiments have suggested that additional resources may in fact have sizable effects on student achievement. The idea that resources matter cannot in itself explain the black-white achievement gap, because most school resources are now fairly equally distributed between blacks and whites. But certain crucial resources, such as teachers with high test scores, are still unequally distributed. And other resources, such as small classes and teachers with high expectations, may help blacks more than whites.<sup>31</sup> The idea that resources matter also suggests that "compensatory" spending on black schools could be valuable, at least if the money were used to cut class size and implement policies that have been shown to help.

This book tries to bring together recent evidence on some of the most controversial and puzzling aspects of the test score debate.<sup>32</sup> Section I examines the role of test bias, heredity, and family background in the black-white gap. Section II looks at how and why the gap has changed over the past generation. Section III examines educational, psychological, and cultural explanations for the gap. Section IV analyzes some of the educational and economic consequences of the gap. The book concludes with a commentary by William Julius Wilson. The rest of the introduction summarizes the book's main findings and then discusses some of their implications.

### Test Bias

Many blacks and some whites believe that all cognitive tests are racially biased. In chapter 2 Christopher Jencks discusses five possible varieties of racial bias in testing. He concludes that two of the five constitute serious problems and that three are probably of minor importance.

29. Steele (1992, 1997).

30. Coleman and others (1966); Mosteller and Moynihan (1972); Jencks and others (1972); and Hanushek (1989).

31. See chapters 8 and 9.

32. The reader who wants a more complete survey should look at Miller (1995); Brody (1992, chap. 10); Snyderman and Rothman (1990); Vernon (1979); and Loehlin, Lindzey, and Spuhler (1975).

*Labeling Bias*

What Jencks calls “labeling bias” arises when a test claims to measure one thing but really measures something else. This is a major problem when tests claim to measure either intelligence or aptitude, because these terms are widely used to describe innate “potential” as well as developed abilities. The notion that intelligence and aptitude are innate seems to be especially salient in discussions of racial differences. Thus, the statement that “blacks are less intelligent than whites” is widely understood as a statement about innate differences. Yet almost all psychologists now agree that intelligence tests measure developed rather than innate abilities, and that people’s developed abilities depend on their environment as well as their genes. Even psychologists who believe that racial differences in test performance are to some extent innate agree that intelligence tests overstate the difference one would observe if blacks and whites grew up in identical environments. Intelligence tests therefore constitute a racially biased estimate of innate ability, which is what nonpsychologists often mean by the word “intelligence.” Test designers cannot eliminate this bias by changing the content of intelligence tests. The only way to eliminate it is to change the tests’ labels so as to emphasize the fact that they measure developed rather than innate skills and abilities.

*Content Bias*

“Content bias” arises when a test contains questions that favor one group over another. Suppose, for example, that black and white children spoke mutually unintelligible versions of English. A test given in white English would then underestimate black children’s skills and vice versa. This kind of content bias does not appear to be a significant problem for the tests discussed in this book. If one takes a standard vocabulary test and winnows out words with unusually large black-white differences, for example, the black-white gap does not shrink much. Likewise, if one compares black children to slightly younger white children, blacks and whites find the same words easy and difficult. Nor is the black-white gap on tests that measure familiarity with the content of American culture consistently larger than the gap on nonverbal tests that do not measure familiarity with any particular culture. Because the racial gap in children’s test performance is not confined to items that measure exposure to white language, culture,

or behavior but is dramatically reduced when black children are raised in white homes, Jencks suggests that it may reflect differences in the way blacks and whites are taught to deal with what they do not know and in the emphasis they put on learning new cognitive skills.

### *Methodological Bias*

Methodological bias arises when we assess mastery of some skill or body of information in a way that underestimates the competence of one group relative to another. Methodological bias would be important if, say, having black rather than white testers changed the relative standing of black and white test takers. That does not appear to be the case. There is some evidence that describing a test in different ways can affect different groups' relative performance, but we do not yet know how general this is.<sup>33</sup>

### *Prediction Bias*

A generation ago many egalitarians argued that using the SAT to screen applicants for selective colleges was unfair to blacks because tests of this kind underestimated black applicants' academic potential. For most colleges, academic potential means undergraduate grades. Almost all colleges have found that when they compare black and white undergraduates who enter with the same SAT scores, blacks earn *lower* grades than whites, not just in their first year but throughout their college careers.<sup>34</sup> Likewise, when firms compare black and white workers with the same test scores, blacks usually get slightly lower ratings from their supervisors and also do a little worse on more objective measures of job performance.<sup>35</sup> In psychological parlance, this means that tests like the SAT do not suffer from "prediction bias."

### *Selection System Bias*

The test score gap between black and white job applicants has traditionally averaged about one standard deviation. When employers do not screen workers, the performance gap is likely to be much smaller—typi-

33. See chapter 11.

34. See chapters 12 and 13.

35. Hartigan and Wigdor (1989); and Wigdor and Green (1991).

cally more like two-fifths of a standard deviation.<sup>36</sup> The reason for this discrepancy is not that blacks perform better than whites with the same test scores. The reason is that test scores explain only 10 to 20 percent of the variation in job performance, and blacks are far less disadvantaged on the noncognitive determinants of job performance than on the cognitive ones.

Because blacks perform no better on the job than whites with similar scores, many people assume that using tests to select workers is racially fair. But if racial fairness means that blacks and whites who could do a job equally well must have an equal chance of getting the job, a selection system that emphasizes test scores is almost always unfair to most blacks (and to everyone else with low test scores). Imagine a company that has 600 applicants for 100 openings. Half the applicants are black and half are white. If the firm hires all applicants as temporary workers and retains those who perform best on the job, and if the performance gap between blacks and whites averages 0.4 standard deviations, about 36 blacks will get permanent jobs. If the firm selects the 100 applicants with the highest scores, about 13 blacks will get permanent jobs.<sup>37</sup> Jencks argues that the first outcome should be our yardstick for defining racial fairness. Using this yardstick, the second system is clearly biased against blacks. In effect, Jencks says, the second system forces blacks to pay for the fact that social scientists have unusually good measures of a trait on which blacks are unusually disadvantaged.

### The Heredity-Environment Controversy

When the U.S. Army launched the world's first large-scale mental testing program in 1917, it found that whites scored substantially higher than

36. A test score gap of 1 SD implies a job performance gap of 0.4 SDs if the correlation between test scores and performance is 0.4 or—more realistically—if the correlation between test scores and performance is 0.25 and the gap between blacks and whites with the same test score averages 0.15 SDs (Hartigan and Wigdor, 1989).

37. In the test score case the black mean is  $-0.5$  SDs and the white mean is  $+0.5$  SDs. If the within-race SDs are equal, they are both  $(1 - 0.52)^{0.5} = 0.866$ . If the cutpoint is set at  $+1$  SD, blacks must be  $1.5/0.866 = 1.73$  SDs above the black mean, whereas whites must be only  $0.5/0.866 = 0.577$  SDs above the white mean. The selection rates are then 4.2 percent for blacks and 28.3 percent for whites. Blacks will therefore get  $4.2/(28.3 + 4.2) = 13$  percent of the jobs. In the case of temporary workers the black mean is  $-0.2$  SDs, the white mean is  $+0.2$  SDs, and the within-race SDs are 0.98, so blacks must be  $1.2/0.98 = 1.22$  SDs above the black mean and whites must be  $0.8/0.98 = 0.816$  SDs above the white mean. The selection rates are then 11.1 percent for blacks and 20.8 percent for whites, so blacks get  $11.1/(11.1 + 20.8) = 36$  percent of jobs.

blacks. Biological determinists immediately cited these findings as evidence that whites had more innate ability than blacks, but cultural determinists quickly challenged this interpretation. By the late 1930s most social scientists seem to have been convinced that either genetic or cultural factors could explain the gap. Neither side had a convincing way of separating the effects of heredity from the effects of culture, so the debate was an empirical standoff.

After 1945 the horrors of the Holocaust made all genetic explanations of human differences politically suspect. Once the U.S. Supreme Court declared *de jure* racial segregation unconstitutional in 1954, genetic explanations of racial differences became doubly suspect because they were identified with southern resistance to desegregation. As a result, environmentalism remained hegemonic throughout the 1960s. Then in 1969 Arthur Jensen published an article in the *Harvard Educational Review* arguing that educational programs for disadvantaged children initiated as part of the War on Poverty had failed, and that the black-white test score gap probably had a substantial genetic component.<sup>38</sup> Jensen's argument went roughly as follows:

—Most of the variation in white IQ scores is genetic.<sup>39</sup>

—No one has advanced a plausible environmental explanation for the black-white gap.

—Therefore it is more reasonable to assume that part of the black-white gap is genetic than to assume it is entirely environmental.

Jensen's article created such a furor that psychologists once again began looking for evidence that bore directly on the question of whether racial differences in test performance were partly innate. Richard Nisbett reviews their findings in chapter 3.

Two small studies have tried to compare genetically similar children raised in black and white families. Elsie Moore found that black children adopted by white parents had IQ scores 13.5 points higher than black

38. Jensen (1969).

39. Based on his estimates of the correlations between twins reared together and apart, adopted siblings, and other kinds of relatives, Jensen suggested that 80 percent of the variance in white children's test scores was traceable to genetic differences. More recent estimates usually put the percentage between 30 and 70 percent (Jencks and others, 1972, appendix A; Plomin and DeFries, 1980; Rao and others, 1982; and Chipuer, Rovine, and Plomin, 1990). We regard 50 percent as a reasonable middle-of-the-road estimate for white children. Recent work (summarized in Plomin and Petrill, 1997) suggests that genes explain more of the variance in adults' scores than in children's scores, perhaps because adults are in a better position to choose environments compatible with their genetic propensities.



children adopted by black parents.<sup>40</sup> Lee Willerman and his colleagues compared children with a black mother and a white father to children with a white mother and a black father. The cleanest comparison is for mixed-race children who lived only with their mother. Mixed-race children who lived with a white mother scored 11 points higher than mixed-race children who lived with a black mother.<sup>41</sup> Since the black-white IQ gap averaged about 15 points at the time these two studies were done, they imply that about four-fifths of that gap was traceable to family-related factors (including schools and neighborhoods).<sup>42</sup>

A better-known study dealt with black and mixed-race children adopted by white parents in Minnesota. The mixed-race children were adopted earlier in life and had higher IQ scores than the children with two black parents. When the 29 black children were first tested, they scored at least ten points higher than the norm for black children, presumably because they had more favorable home environments than most black children.<sup>43</sup> When these children were retested in their late teens or twenties, their IQ scores had dropped and were no longer very different from those of Northern

40. Moore (1986) studied 23 white and 23 black families. The sample included 20 biracial children and 26 all-black children. In black families the 6 biracial children scored 2.8 points higher than the 17 all-black children. In white families the 14 biracial children scored 1.5 points lower than the 9 all-black children.

41. Willerman, Naylor, and Myriantopoulos (1974) studied 129 four-year-olds, 101 of whom were raised by a white mother and 28 of whom were raised by a black mother. Among the married mothers the 50 children raised by a white mother and a black father had mean IQs of 104.7, while the 17 children raised by a black mother and a white father had mean IQs of 96.4. Among single mothers, who provide a cleaner comparison of the effects of growing up in a "white" rather than a "black" environment, the 51 children raised by a white mother had mean IQs of 99, while the 11 children raised by a black mother had mean IQs of 88. The mixed-race children raised by black and white mothers presumably had about the same mixture of "African" and "European" genes, but the parents may have been subject to different forms of genetic selection relative to others in their racial group.

42. Children raised by, say, black and white single mothers usually grow up in different neighborhoods and attend different schools. Nonetheless, black children raised by white parents are unlikely to have exactly the same overall environment as white children raised by the same parents. We do not know how white parents treat their black and white children, but there could be important differences favoring either group. We do know that teachers often treat black and white children somewhat differently (see chapter 8), and strangers are even more likely to do so. In addition, black and white children see themselves somewhat differently and may choose different environments as a result.

43. Most adoptees were first tested between the ages of four and sixteen. Weinberg, Scarr, and Waldman (1992, table 2) report data indicating that at the time of initial testing IQ scores averaged 97 for adoptees known to have had two black parents ( $N = 29$ ), 109 for adoptees known to have had one black and one white parent ( $N = 68$ ), and 109 for adoptees with a black mother and a father of unknown race ( $N = 33$ ). The IQ gap between all-black and mixed-race children was much larger in the Minnesota study than in other studies (Moore, 1986; see also Flynn, 1980, on Eyferth).

blacks raised in black families.<sup>44</sup> The most obvious explanation for this drop is that the adoptees had moved out of their white adoptive parents' homes into less favorable environments.<sup>45</sup> But because the study did not cover black or mixed-race children adopted by black parents, it does not seem to us to provide strong evidence on either side of the heredity-environment debate.

### *Racially Mixed Children*

Race is not a well-defined biological category. It is a social category, whose biological correlates vary geographically and historically. America has traditionally classified people as black using the "one drop" rule, under which anyone with known black ancestors is black. As a result, people are often treated as black even though they have a lot of European ancestors. If blacks with a lot of European ancestors had the same test scores as those with no European ancestors, we could safely conclude that the black-white test score gap was a by-product of social classification rather than heredity. But when we find that light-skinned blacks score higher than dark-skinned blacks, we cannot rule out the possibility that this difference is environmental. Light skin has traditionally been a social asset for black Americans, and the correlation between light skin and test performance could reflect this fact. To get around this problem, we need less visible genetic markers. Two studies have used blood markers to estimate the percentage of Europeans in a black child's family tree. Neither study found a correlation between the number of "European" blood markers and IQ.<sup>46</sup>

44. Weinberg, Scarr, and Waldman (1992) report data suggesting that when the Minnesota adoptees were retested, the IQ means were 89 for black adoptees (N = 21), 99 for mixed-race adoptees (N = 55), and 99 for those whose father's race was unknown (N = 25). Adjusting for changes in test norms makes the decline about 2 points larger (Waldman, personal communication, August 6, 1997).

45. It is also conceivable that genetic differences between blacks and whites become more important in adolescence. McCartney and others' (1992) meta-analysis of twin studies suggests that the relative importance of home environment falls in adolescence and that the relative importance of genes increases. Thus if black adoptees had average test scores prior to adolescence because the negative effects of their genes were offset by the positive effects of their adoptive homes, and if their home environment exerted less influence after they reached adolescence, their test scores might fall.

46. Loehlin, Vandenberg, and Osborne (1973) and Scarr and others (1977). These studies would be more convincing if blood group classification were more reliable. With today's technology it should be possible to classify children more accurately along the European-versus-African ancestry continuum, but conducting such a study would be politically difficult.

Although racially mixed children are culturally black in America, and are almost always raised by black parents in black communities, this is not true everywhere. Klaus Eyferth studied the illegitimate children of black and white soldiers stationed in Germany as part of the army of occupation after World War II. All these children were raised by their German mothers.<sup>47</sup> There was considerable prejudice against blacks in Germany at the time, and any child of a German mother who looked black was also presumed to be illegitimate, which carried additional stigma. But mixed-race German children did not attend predominantly black schools, live in black neighborhoods, or (presumably) have predominantly black (or mixed-race) friends. When Eyferth gave these children a German version of the Wechsler IQ test, children with black fathers and white fathers had almost identical scores.<sup>48</sup>

Taken in isolation, none of these studies would carry much weight. The samples are small, and the comparisons could be distorted by unmeasured genetic or environmental influences. But Nisbett argues that their consistency gives them far more weight that they would have if taken one by one.<sup>49</sup> We agree. We read these studies as supporting three tentative conclusions:

—When “black” genes are not visible to the naked eye and are not associated with membership in a black community, they do not have much effect on young children’s test scores.

—Growing up in an African-American rather than a European-American family substantially reduces a young child’s test performance.

—When black Americans raised in white families reach adolescence, their test scores fall.

47. The most detailed summary of this study in English is found in Flynn (1980). Eyferth sampled roughly 5 percent ( $N = 181$ ) of the children known to have been fathered by black French and American soldiers between 1945 and 1953. He then constructed a matched sample of 83 children fathered by white soldiers. The two samples were matched only on characteristics of the mother and location, not characteristics of the father, which were largely unknown. Eighty percent of the black fathers were American and 20 percent were French Africans. Flynn reports that in the U.S. army of occupation the black-white gap on the Army General Classification Test (a predecessor of the AFQT) was about four-fifths that in the general population.

48. The means were 96.5 for mixed-race children and 97.2 for all-white children. The sampling error of this difference is about 4 points.

49. Although the published results are quite consistent, there is some risk that unpublished studies might show a different pattern, especially if liberal authors engage in self-censorship or liberal reviewers have higher scientific standards for papers that they see as dangerous. But while political correctness could have posed an important obstacle to publication in sociology and education journals, a number of psychology journals regularly publish papers suggesting that the black-white gap is partly innate.

These studies do not prove that blacks and whites would have exactly the same test scores if they were raised in the same environment and treated the same way. But we find it hard to see how anyone reading these studies with an open mind could conclude that innate ability played a *large* role in the black-white gap.<sup>50</sup>

### Effects of Family Background

Chapter 4, by Meredith Phillips, Jeanne Brooks-Gunn, Greg Duncan, Pamela Klebanov, and Jonathan Crane tries to estimate the effect of specific family characteristics on young children's test scores. This is not easy. Hundreds of different family characteristics correlate with children's test performance. Disentangling their effects is a statistical nightmare. Almost any family characteristic can also serve as a proxy for a child's genes. We know, for example, that a mother's genes affect her test scores and that her test scores affect her educational attainment. Thus when we compare children whose mothers finished college to children whose mothers only finished high school, the two groups' vocabulary scores can differ for genetic as well as environmental reasons. Even when a child is adopted, moreover, the way the adoptive parents treat the child may depend on the child's genes. Parents read more to children who seem to enjoy it, for example, and whether children enjoy being read to may well depend partly on their genes.<sup>51</sup>

The best solution to such problems is to conduct experiments. In the 1970s, for example, the federal government conducted a series of "negative income tax" experiments that increased the cash income of randomly se-

50. Skeptics may wonder whether other experts read this literature the way we do. That question is not easy to answer. Snyderman and Rothman (1987) asked a sample of over 1,000 psychologists, sociologists, and educational researchers, "Which of the following best characterizes your opinion of the heritability of the black-white differences in IQ?" Of the 661 "experts" who returned a questionnaire, 14 percent declined to answer this particular question, 24 percent said the data were insufficient to support a reasonable opinion, 1 percent thought the gap was "due entirely to genetic variation," 15 percent thought it was "due entirely to environmental variation," and 45 percent thought it was "a product of both genetic and environmental variation." It is not clear how many of those who gave the "both" response would accept our conclusion that genes do not play a *large* role in the black-white gap. Nor is it clear how many of Snyderman and Rothman's respondents had read the research that Nisbett reviews.

51. For a general review see Plomin and Bergeman (1991). For data on parental treatment of adopted children and twins see Braungart, Fulker, and Plomin (1992).

lected low-income families. These experiments did not last very long, the samples getting any given “treatment” were small, and the results were poorly reported, so it is hard to know exactly what happened. Short-term income increases did not have statistically reliable effects on low-income children’s test scores, but that does not mean their true effect was zero.<sup>52</sup> As far as we know, these are the only randomized experiments that have altered families’ socioeconomic characteristics and then measured the effect on children’s test scores.

In theory, we can also separate the effects of parents’ socioeconomic status from the effects of their genes by studying adopted children. But because adoption agencies try to screen out “unsuitable” parents, the range of environments in adoptive homes is usually restricted. The adoptive samples for which we have data are also small. Thus while parental SES does not predict adopted children’s IQ scores as well as it predicts natural children’s IQ scores, the data on adopted children are not likely to persuade skeptics.<sup>53</sup>

Anyone who wants to estimate the impact of specific family characteristics on children’s test scores must therefore rely heavily on surveys of children raised by their natural parents. The best source of such data is the National Longitudinal Survey of Youth (CNLSY) that Phillips and her colleagues use in chapter 4. Black five- and six-year-olds in their sample scored about 16 points (one standard deviation) below whites on the Peabody Picture Vocabulary Test (PPVT). Traditional measures of educational and economic inequality do not explain much of this gap. Measures of a mother’s socioeconomic position when she was growing up and measures of her current parenting practices explain considerably more.

52. See the summary in Mayer (1997).

53. The only convincing way to deal with unmeasured restrictions in the range of adoptive children’s home environments is to compare the regression of IQ on parental SES for adopted and natural children in the same families. Scarr and Weinberg (1978) report much larger correlations for natural than adopted children in a postadolescent sample. Comparisons involving younger samples are ambiguous. The IQ-SES correlations for adopted children in Scarr and Weinberg (1976) do not differ in any consistent way from the correlations for natural children in (more or less) the same families (Scarr and Weinberg, 1978). The IQ-SES correlations for natural and adopted children in Loehlin, Horn, and Willerman (1989) also seem inconclusive. For early studies without a comparison group of natural children raised in the same family see Jencks and others (1972, p. 80). For a small recent study of separated identical twins with an unusually large variance for parental education see Bouchard and others (1990).

### *Parental Schooling*

Early in the twentieth century white parents typically completed two or three more years of school than blacks.<sup>54</sup> By 1991 the gap between black and white mothers with children in first grade had fallen to 0.8 years.<sup>55</sup> Many observers have suggested that this change played a significant role in reducing the black-white gap in children's test scores.<sup>56</sup> But if parental schooling correlates with children's test scores mainly because it is a proxy for parental genes, changing the distribution of schooling will not affect the distribution of test scores, either for individuals or for groups.

When Phillips and her colleagues control a mother's family background, the estimated effect of an extra year of school on her child's PPVT score falls from 1.73 to 1.15 points.<sup>57</sup> When they also control the mother's AFQT score (their proxy for her cognitive genotype), the effect falls to somewhere between 0.5 and 0.6 points.<sup>58</sup> This suggests that a two-year reduction in the black-white education gap among mothers would cut the PPVT gap by about a point. Of course, if the schooling gap narrowed because black and white parents had grown up in more similar homes or had more similar test scores, the predicted reduction in the PPVT gap between black and white children would be larger.<sup>59</sup> The CNLSY suggests that cutting the

54. Hauser and Featherman (1976) show, for example, that the fathers of black men born between 1907 and 1916 had 2.6 years less schooling than the fathers of white men born in these years. We could not find comparable data on mothers.

55. This estimate comes from the Prospects sample that Phillips, Crouse, and Ralph describe in chapter 7. The racial disparity in mother's education is smaller in CNLSY than in Prospects (0.2 versus 0.8 years), at least partly because the CNLSY data that Phillips and her colleagues analyzed underrepresent older mothers, who are disproportionately white and highly educated.

56. See chapters 5 and 6, as well as Armor (1992).

57. See Chapter 4, table 4-3.

58. Phillips and her colleagues include some mothers whose test scores were measured after they had had different amounts of schooling. For these mothers, differences in schooling contribute to differences in AFQT scores as well as the other way round. Winship and Korenman (1997) estimate that an extra year of schooling raises test scores by an average of 0.18 SDs. Phillips and her colleagues estimate that with maternal education controlled a 1 SD increase in a mother's AFQT score raises her child's test score by 0.23 SDs. Since the coefficient of mother's schooling with AFQT controlled is 0.52, the sum of the direct and indirect effects of an extra year of maternal schooling on a child's PPVT score is  $0.52 + (0.18)(0.23) = 0.56$  points. This estimate assumes that a mother's IQ genotype has no partial correlation with her educational attainment once we control her AFQT score.

59. Black and white CNLSY parents also attended schools of differing quality. The CNLSY shows that the percentage of whites in a mother's high school had a significant positive effect on her children's PPVT score, independent of her own race, years of schooling, and AFQT score. Teacher turnover also had a significant negative effect.

schooling gap between black and white fathers would have a smaller effect than cutting the gap between black and white mothers. But this may not be true for older children, for whom the effects of mother's and father's education are roughly equal.<sup>60</sup>

### *Income Effects*

White CNLSY parents reported 73 percent more income than their black counterparts. When Phillips and her colleagues compared black and white children whose families had had the same average annual income since the child was born, the PPVT gap narrowed by 2.4 points. But once again it does not follow that raising all black families' incomes by 73 percent would raise their children's PPVT scores by 2.4 points. To estimate the effect of increasing black parents' incomes without changing the traits that cause the current income gap, we need to know how much parental income affects children's test scores when we compare parents with the same family background, test scores, and schooling. These controls cut the estimated effect of parental income on PPVT scores by about three-fifths. Even this estimate is probably on the high side, because it does not control either the father's test scores or his family background. Thus, the CNLSY suggests that eliminating black-white income differences would cut the PPVT gap by less than 1 point.<sup>61</sup> Eliminating the causes of the black-white income gap might, of course, have a much larger effect. Racial disparities in parental wealth have almost no effect on children's test scores once Phillips and her colleagues control income, schooling, and the mother's test scores.

### *Single-Parent Families*

Once Phillips and her colleagues hold constant the mother's family background, educational attainment, and test scores, children who have grown up in an intact family score no higher on the PPVT than children

60. See table 5B-3 of chapter 5.

61. This estimate is based on the fact that (a) equation 4 in table 4-3 of chapter 4 shows that controlling the natural log of mean family income since birth reduces the test score gap from 16.27 to 13.89 points, and (b) equations 5 and 7 in the same table show that controlling parental schooling, mother's AFQT, maternal grandparents' SES, and a set of location variables lowers the coefficients of the parental income dummies by an average of 60 percent. The estimated reduction in the black-white PPVT gap due to income per se is thus about  $(16.27 - 13.89)(0.40) = 0.95$  points, or 6 percent of the initial gap.

from single-parent families. Other studies find slightly larger effects, but the effects are never large enough to be of any substantive importance.<sup>62</sup>

### *Parenting Strategies*

Knowing parents' education and income tells us something about how they raise their children, but not much. The CNLSY tried to measure parenting practices directly, both by asking parents what they did and by asking interviewers how mothers treated their children during the interview. Parenting practices appear to have a sizable impact on children's test scores. Even with parental education, family income, and the mother's AFQT scores controlled, racial differences in parenting practices account for between a fifth and a quarter of the racial gap on the PPVT. This suggests that changes in parenting practices might do more to reduce the black-white test score gap than changes in parents' educational attainment or income. We cannot be sure how large these effects would be, however, because the way parents treat their children is a proxy for all kinds of unmeasured parental characteristics, as well as for the child's own genes.<sup>63</sup>

### *Grandparents*

Upwardly mobile parents often raise their children the way they themselves were raised. Phillips and her colleagues find that racial differences in parenting practices are partly traceable to the fact that even when black and white parents have the same test scores, educational attainment, income, wealth, and number of children, black parents are likely to have grown up in less advantaged households. Phillips and her colleagues also find that

62. Using the 1980 High School and Beyond data, Mulkey, Crain, and Harrington (1992) found that even without controlling parental test scores, controls for parental schooling and race reduced the estimated effect of living with a single mother on tenth-grade vocabulary scores to 0.06 standard deviations. McLanahan and Sandefur (1994, p. 44) report that living with a single mother reduced quartile rank (coded 1-4) by 0.11 points, which is equivalent to a reduction of  $(0.11)(25) = 3$  percentile points. This is consistent with Mulkey, Crain, and Harrington's findings.

63. On genetic determinants of parenting practices see Plomin, DeFries, and Fulker (1988); Braungart, Fulker, and Plomin (1992); and Plomin (1994). Although these data suggest that children's genes have some effect on parenting practices, the effects appear to be small and the standard errors of the estimates are large.



this can lower black children's test scores. In other words, it can take more than one generation for successful families to adopt the "middle-class" parenting practices that seem most likely to increase children's cognitive skills.

### Changes in the Black-White Gap

Chapter 5, by Larry Hedges and Amy Nowell, analyzes changes since 1965 in the black-white test score gap at the end of high school. As figure 1-2 showed, this gap has narrowed substantially. Nonetheless, racial disparities are still very large, especially near the top of the distribution. In both 1982 and 1992 white high school seniors were about ten times more likely than black seniors to score in the top 5 percent of the national distribution on a composite measure of academic skills. These test score disparities near the top of the distribution put blacks at a severe disadvantage in competition for places at selective colleges and professional schools. They are also likely to have a major impact on the racial composition of the nation's professional and managerial elite.

David Grissmer, Ann Flanagan, and Stephanie Williamson also examine changes in the black-white gap in chapter 6. Unlike Hedges and Nowell, they analyze trends in black and white achievement separately, asking what could have caused the changes they document for each group. This approach focuses attention on a puzzle that neither they nor anyone else has been able to resolve fully. It is easy to identify changes in both families and schools that might account for the improvement in black achievement. Increased parental education, smaller families, smaller classes, and rising enrollment in academically demanding courses are all plausible candidates. But these changes have all affected whites as well as blacks. The changes were smaller for whites than for blacks, but they were still substantial. If changes of this kind raised black scores, they should also have raised white scores. Yet as figure 1-2 indicates, white seventeen-year-olds' reading and math scores rose by only 0.1 standard deviations between the early 1970s and 1996.

One way to solve this puzzle is to assume that improvements in schooling were more valuable to blacks than to whites. But while there is some evidence that blacks benefit more than whites from small classes and highly educated mothers, such advantages always confer some benefit on whites as well. Grissmer and his colleagues therefore looked for changes that might

plausibly be expected to raise blacks' scores without affecting whites' scores. School desegregation in the South is one example. Desegregation seems to have pushed up southern blacks' scores a little without affecting whites either way. Grissmer and his colleagues also speculate that the civil rights movement, the War on Poverty, and affirmative action may have changed black parents' and students' attitude toward school achievement without affecting whites, but this hypothesis is hard to test in a rigorous way.

The authors also suggest that the decline in black teenagers' reading scores after 1988 could be related to the surge in violence among black teenagers that began in the late 1980s. All else equal, the surge in violence should also have depressed black adolescents' math scores.<sup>64</sup> But the negative effect of violence on math scores may have been offset by the fact that more black adolescents were studying geometry and second-year algebra. Among non-Hispanic whites, who were also taking more demanding math courses and were not murdering one another at a higher rate, math scores have risen steadily since 1982, while reading scores have remained flat.

### *Does the Gap Widen with Age?*

Discussions of how schooling affects the black-white achievement gap have long been bedeviled by disagreement about how the gap should be measured. Those who think the gap widens with age almost always describe achievement in terms of grade levels or years. When a first-grader scores at the 16th percentile of the first-grade distribution, for example, she has a vocabulary roughly comparable to the average kindergartner. One way to describe this situation is to say that she is "one year below grade level." When a seventh-grader scores at the 16th percentile of the seventh-grade distribution, she has a vocabulary roughly comparable to that of the average fifth-grader, so we can say she is now "two years below grade level." But it is not clear that a child who consistently scores at the 16th percentile is really farther behind at age twelve than at age six. That makes us reluctant to discuss the black-white gap in these terms, and none of the contributors to this volume does so.

64. Precise comparisons between trends in reading and math achievement are impossible before 1990 because NAEP did not give its reading and math tests in the same years. We have no data on math achievement in 1988, when black reading achievement peaked. Thus, we do not know whether black seventeen-year-olds' math achievement deteriorated as much as their reading achievement between 1988 and 1990. There was no clear trend in either reading or math achievement among black seventeen-year-olds from 1990 to 1996.

Instead, they focus on where blacks fall in the national distribution at a given age. They usually describe their findings in terms of “standard deviations.” If the average black scores at the 16th percentile, the average white scores at the 50th percentile, and the distribution is normal, the black-white gap is one standard deviation. Sometimes contributors simply report the average black child’s percentile rank in the overall distribution of scores. If every black child falls at the same percentile of the overall distribution at age eighteen as at age six, what psychometricians call “age standardized” measures will show no change in the gap. Chapter 7, by Meredith Phillips, James Crouse, and John Ralph, shows that the age-standardized black-white gap hardly changes for reading between first and twelfth grades. The math gap increases a little. The vocabulary gap appears to increase more, but the vocabulary data are not as good as the reading or math data, so it is hard to be sure.<sup>65</sup>

Such findings have persuaded some social scientists that schools do not contribute much to the black-white gap. This conclusion strikes us as premature. If stability were the natural order of things and did not need to be explained, the fact that blacks typically score at about the same percentile in first and twelfth grades would be evidence that schools did not contribute to the gap. In human development, however, change is at least as common as stability. That means we have to explain stability as well as change.

Consider a simple example. Suppose we give children a vocabulary test at the end of kindergarten. Twelve years later we retest both blacks and whites who scored at the 16th percentile of the white distribution in kindergarten. Since we have selected white children with unusually low first-grade scores, most are likely to have come from homes that were considerably worse than the average white child’s home. But home environments change in unpredictable ways between first and twelfth grades. Mothers who are constantly irritated by small children sometimes do much better with older children. Fathers who drink too much sometimes go on the wagon. Unstable jobs and living arrangements sometimes stabilize. Children’s envi-

65. Both Entwisle, Alexander, and Olson’s (1997) Baltimore data and the CNLSY show that the standardized black-white gap on reading and math tests is quite small for five- and six-year-olds. In the case of the CNLSY, this may be because the reading and math tests for five- and six-year-olds have poor psychometric properties. The PPVT-R, which is less sensitive to formal instruction, shows less increase in the standardized black-white gap with age. Even tests that are sensitive to formal instruction, like math tests, do not show much increase in the standardized black-white gap after the first few years of school.

ronments also expand as they get older. Some children of neglectful parents encounter dedicated teachers. Some children from semiliterate homes make friends with children who read a lot, while others discover libraries. Because these changes have a large random element, children whose homes were unusually bad during their first five years of life tend to have somewhat better environments later in life.<sup>66</sup> This process is often called “regression to the mean,” but that is only a label, not an explanation. Readers may find the reasons for regression to the mean easier to understand if they focus on the child whose home ranked at the very bottom at the end of kindergarten. Twelve years later, this child’s home is still likely to be relatively bad, but it is unlikely to be the very worst.

Because every child’s environment changes, white children who started school at the 16th percentile seldom score at exactly the 16th percentile in twelfth grade. Nor do their scores spread out in a completely random way. Some score below the 16th percentile in twelfth grade, but most score above it. This happens because the luck of repeated environmental draws moves most children with low initial scores closer to the average.<sup>67</sup> A reasonable guess is that the average white child who starts school scoring at the 16th percentile will finish school scoring at about the 27th percentile.<sup>68</sup>

Now consider the black children who started school scoring at the 16th percentile of the white distribution. If the black-white gap does not change with age and is always one standard deviation, blacks who score at the 16th percentile in first grade will still score at the 16th percentile in twelfth grade. Black and white children who start out at the 16th percentile will therefore end up with quite different scores in twelfth grade, with the average white at the 27th percentile and the average black at the

66. The genes that affect test performance may also have different effects at different ages, which would produce the same regression effects.

67. Readers should also bear in mind that 16 percent of children always score below the 16th percentile, regardless of the children’s age. The fact that children who start near the bottom tend to move up therefore has to be offset by the fact that other children who started with higher scores move down to fill the bottom rungs of the ladder. Since 84 percent of all children start out above the 16th percentile, only a small minority of this large group needs to slip below the 16th percentile to replace those who move up.

68. This estimate assumes that the observed correlation between first- and twelfth-grade verbal scores is 0.52. The estimate comes from a sample of 753 children in Boulder, Colorado, who took the California Test of Mental Maturity in first or second grade and took the Lorge-Thorndike intelligence test in eleventh grade (Hopkins and Bracht, 1975). Assuming an average test-retest reliability of 0.85 for verbal IQ at these ages, the true correlation is  $0.52/0.85 = 0.61$ . The estimate is meant to be heuristic, not exact.

16th percentile. Something must cause this difference. That “something” could involve genes, home environments, school environments, or other factors.<sup>69</sup>

Phillips and her colleagues’ best guess is that if we followed representative samples of black and white children from first to twelfth grade we would find that about half the black-white gap in twelfth-grade reading and math scores was explained (in a statistical sense) by differences in their first-grade scores. The other half would then be attributable to things that happened between the ages of six and eighteen. But this estimate need not mean that schools cause half the difference between black and white twelfth-grade scores. A number of studies show, for example, that the test score gap between economically advantaged and disadvantaged children widens more over the summer, when children are not in school, than it does during the winter.<sup>70</sup> This suggests that racial differences in how much children learn between the ages of six and eighteen depend partly on what happens outside the schools. Nonetheless, schools are society’s main instrument for altering the distribution of cognitive skills. Their role in widening or narrowing the black-white gap therefore requires strict scrutiny.

### What Can Schools Do?

Instead of asking whether schools cause the black-white test score gap, Ronald Ferguson asks what they can do to reduce it. In chapter 8 he deals with the vexed question of teacher expectations. His survey of the evidence leads him to five conclusions.

- Teachers have lower expectations for blacks than for whites.
- Teachers’ expectations have more impact on black students’ performance than on white students’ performance.
- Teachers expect less of blacks than of whites because black students’ past performance and behavior have been worse. Ferguson finds no evi-

69. Some readers may think it is obvious that whites should regress to the white mean and that blacks should regress to the black mean. These readers should reframe the puzzle in terms of a simple regression model. Let  $T$  denote a test score, let the subscripts 1 and 2 denote time, let  $W$  be a dichotomous variable with a value of 1 for whites and 0 for nonwhites, and let  $e$  be an error term. Then  $T_2 = B_0 + B_1 T_1 + B_w W + e$ . The question is why  $B_w$  should be positive even after correcting  $T_1$  for measurement error.

70. On summer versus winter learning, see Heyns (1978, 1987) and Entwisle and Alexander (1992) and the sources they cite.

dence that teachers' expectations differ by race when they are asked to assess children who have performed equally well and behaved equally well in the past.

—By basing their expectations on children's past performance and behavior, teachers perpetuate racial disparities in achievement.

—Exhorting teachers to have more faith in black children's potential is unlikely to change their expectations. But professional development programs in which teachers actually see disadvantaged black children performing at a high level can make a difference.

In chapter 9 Ferguson reviews school policies that might help reduce the black-white test score gap. A substantial number of randomized experiments suggest that smaller classes raise test scores.<sup>71</sup> The largest of these studies is the Tennessee class-size experiment, which covered 6,572 children in 76 schools. It found that cutting class size by a third between kindergarten and third grade (from about 23 to about 15) raised both reading and math scores by about a third of a standard deviation for blacks and by a sixth of a standard deviation for whites. After children moved to larger classes in fourth grade, the benefits associated with having been in smaller classes began to diminish, especially for blacks. But the benefits were still sizable at the end of the seventh grade, which is the last year for which we currently have data.<sup>72</sup>

History is never as tidy as a randomized experiment, but the historical record seems consistent with the hypothesis that reducing class size raises reading scores. Averaged across all grade levels, the pupil-teacher ratio fell from 26:1 in the early 1960s to 17:1 in the early 1990s. About half the extra teachers were used to cut the number of students in regular classrooms. The size of the average class therefore fell from 29 in 1961 to 24 in 1991.<sup>73</sup> Ferguson finds that changes in the pupil-teacher ratio predict changes in the black-white reading gap among NAEP nine-year-olds quite well.<sup>74</sup> The pupil-teacher ratio fell substantially in the 1970s. Both black and white nine-year-olds' reading scores rose during this period, but blacks' scores rose more, which is what the Tennessee results would predict. Unfortunately, nine-year-olds' math scores did not rise in tandem with their reading scores. Black nine-year-olds' math scores rose only a little during

71. Glass and others (1982).

72. Mosteller (1995). For a far more skeptical view see Hanushek (1998).

73. Lewitt and Baker (1997).

74. See also chapter 6.

the 1970s, and white nine-year-olds' math scores hardly changed at all.<sup>75</sup> Perhaps the impact of smaller classes on math scores was offset by other countervailing influences, such as a less demanding curriculum.

Counting the number of teachers in a school is easier than measuring their competence. Ferguson argues that a teacher's test score is the best readily available indicator of the teacher's ability to raise children's test scores. No one has done randomized experiments to see whether hiring teachers with higher test scores improves student achievement, but a large body of non-experimental research by Ferguson and others suggests that high-scoring teachers are more effective. Ferguson therefore concludes that using competency exams to screen out low-scoring teachers will help children in the long run. Screening out teachers with low scores should benefit blacks even more than whites, because black children are now considerably more likely than whites to have teachers with low scores. But competency exams will not do much to raise students' test scores unless the passing score is high enough to eliminate a substantial fraction of potential teachers. At present, few states or school districts set the cut point on competency tests very high.

Unfortunately, raising the passing score on teacher competency exams will also reduce the percentage of blacks who qualify for teaching jobs. This is a major political problem. It may also be a pedagogic problem, since Ferguson finds some evidence that black children learn more from black teachers. But this evidence is far weaker and less consistent than the evidence that teachers with high test scores will raise black children's scores.

School desegregation can also raise black children's achievement under some circumstances. The findings of David Grissmer and his colleagues strongly suggest that desegregation raised southern black nine-year-olds' test scores in the 1970s. In a study covering the early 1990s, Meredith Phillips also found that attending a whiter school probably had a positive effect on black students' reading scores in the early grades. But she found less evidence that the racial mix of middle schools and high schools affected reading scores. Nor did a school's racial mix have a consistent influence on black students' math scores.<sup>76</sup> Since racially mixed schools have higher-scoring teachers, and since we often assume that teachers in racially mixed schools have higher expectations for their students, this finding is puzzling.

75. Both black and white nine-year-olds' math scores did rise after 1982, but class size has not changed much since the late 1970s.

76. Phillips (1997).

One common hypothesis about why black children derive such modest benefits from attending what seem to be better schools is that racially mixed schools track black students into classrooms where their classmates are mostly black and their teachers expect very little. Ability grouping is obviously a contentious issue even in racially homogeneous schools. Students in low-ability classes usually cover less material than those in mixed-ability classes, but covering less material does not necessarily mean that students learn less. Ability grouping is supposed to increase the odds that slower learners get instruction appropriate to their skills. If that happened, slow learners could be better off.

Public debates about ability grouping seldom pay much attention to empirical evidence about its impact, perhaps because the evidence does not fit either side's preconceptions. Ferguson's review suggests that assigning children to separate classrooms on the basis of their presumed ability has no significant effect, positive or negative, on the test scores of children in the middle or bottom third of the distribution. This kind of grouping may help high-scoring students a little, but if so the effect is very small.<sup>77</sup> Assigning students to heterogeneous classrooms and then grouping them by ability *within* a heterogeneous classroom appears to increase math achievement for all children, but there have been only a handful of randomized experiments with this kind of grouping, so it is hard to be sure.<sup>78</sup> The impact of ability grouping on reading skills is apparently unknown. Research on this subject is now mostly qualitative. Only one experimental study of ability grouping and test performance has been conducted since 1974.

At the high school level, black and white students are almost equally likely to say they are in the college preparatory program, but white students are substantially more likely to take academically demanding classes, such as second-year algebra, geometry, or Advanced Placement courses of various kinds. Ferguson finds that class assignments depend on students' pre-

77. In a review limited to randomized experiments, Mosteller, Light, and Sachs (1996) found ten studies covering 2,641 students that compared heterogeneous classrooms to classrooms tracked on the basis of estimated ability or past achievement ("XYZ grouping"). Averaged across all students, the mean difference in achievement between the two methods was exactly zero. There was some evidence that the least skilled students learned less and that the most skilled students learned more under XYZ grouping, but this difference was not statistically significant.

78. Mosteller, Light, and Sachs (1996) found only three experimental studies that compared the effect of ability grouping within heterogeneous classrooms with the effect of ungrouped heterogeneous classrooms.



vious grades, test scores, and socioeconomic background, but not on their race per se. We do not know *how* socioeconomic background affects course assignments. High-SES parents may encourage their children to take more demanding courses even when their children's test scores or grades are marginal. High-SES parents may also pressure schools to let their children take such courses. It would be a mistake to fault schools for letting high-SES students take demanding courses, however. Instead, we should probably fault schools for not pushing more low-SES students into such courses.

### Fear of "Acting White"

An influential 1986 article by Signithia Fordham and John Ogbu drew attention to the fact that academically successful black adolescents often said their classmates disparaged them for "acting white." Some black students also reported that they had stopped working hard in order to avoid such taunts. Since 1986 many successful black adults have also reported such experiences. So have other ethnographers. Indeed, this explanation for blacks' academic problems has become part of American folklore and is often repeated in the mass media. Philip Cook and Jens Ludwig were apparently the first scholars to test such claims quantitatively with national data.<sup>79</sup> Chapter 10 describes their findings.

If black students were less committed to academic success than white students, we would expect blacks to do less homework and skip school more often. Cook and Ludwig find that few students, black or white, do much homework. The median black, like the median white, spends between two and four hours a week doing homework outside school. The hardest-working whites do more homework than the hardest-working blacks, but even for these students the difference is not huge: 14 percent of white tenth-graders report spending more than ten hours a week on homework, compared with 10 percent of blacks. When it comes to skipping school, racial differences are negligible.

Cook and Ludwig also investigate the social costs of academic success for blacks and whites. They do not challenge Fordham and Ogbu's claim that working hard or getting good grades can lead to charges of racial disloyalty. But white students who work hard are also taunted as "nerds" and

79. For a more recent analysis along the same lines see Ainsworth-Darnell and Downey (forthcoming)

“geeks.” Black students’ fear of “acting white” can only exacerbate the black-white test score gap if academic success has higher social costs for blacks than for whites, or if blacks are more responsive than whites to such social costs. Since Fordham and Ogbu studied an all-black high school, they could not investigate these issues.

In 1990 the National Education Longitudinal Survey (NELS) asked tenth-graders whether they had been physically threatened at school, whether they often felt put down by other students in their classes, whether other students thought of them as popular, whether other students thought of them as part of the “leading crowd,” and whether they were popular with the opposite sex. Cook and Ludwig use these measures to assess the social consequences of getting A’s in math and of being a member of the student honor society. Their analysis yields two findings.

—Getting A’s in math is almost unrelated to being threatened, feeling put down by other students, and feeling unpopular. Honor society members appear to feel *less* threatened and *more* popular than other students.

—The social costs and benefits of academic success are about the same for blacks and whites. Where racial differences do arise, blacks usually benefit slightly more than whites from academic success.

Cook and Ludwig’s findings suggest that while academic success can have social costs, it also has benefits. If schools can reduce the costs or increase the benefits, student achievement is likely to improve. Judging by Cook and Ludwig’s findings, however, the improvement would be as large for whites as for blacks. If that were the case, the racial disparity in academic achievement would not change much.

Ferguson challenges these conclusions in his comment on Cook and Ludwig’s chapter. His most important criticism is that they are asking the wrong question. He agrees that fear of acting white probably plays a minor role in *creating* the black-white test score gap that we observe in American high schools, but it may nonetheless be an important obstacle to *reducing* the gap. Ferguson compares American high schools to a mile-long race in which two competitors are nearing the end of the course. They are separated by forty yards. Both are jogging. Neither is out of breath. Observing this scene, we could ask why the black runner was behind. But we could also ask why the black runner was not trying to catch up. The “acting white” hypothesis does not seem to explain why black high school students are behind their white classmates. But it may well explain why they are not making more effort to catch up.

### “Stereotype Threat”

Claude Steele and Joshua Aronson argue in chapter 11 that academically successful blacks worry about the possibility that getting a low score on a test will confirm the stereotype that blacks are not academically talented. This kind of anxiety, they argue, can actually impair successful blacks' performance. They ran a series of ingenious experiments in which a white experimenter gave Stanford undergraduates a test composed of difficult verbal questions from the Graduate Record Exam. Black students made substantially more mistakes when they were asked to record their race before taking the test. Blacks also made more mistakes when they were told that the test measured “verbal reasoning ability” than when they were told that the study focused on “psychological factors involved in solving verbal problems.” White students' performance did not depend on the way the experimenter described the test or on whether they recorded their race.

Steele and Aronson's findings strongly suggest that anxiety about racial stereotypes and intellectual competence can sometimes depress able black students' test performance. But their findings do not tell us how general this phenomenon is. Steele believes that what he calls “stereotype threat” is mainly a problem for blacks who have an emotional investment in seeing themselves as good students.<sup>80</sup> He also believes that it helps explain why so many black students “disidentify” with school.<sup>81</sup> We do not yet know whether stereotype threat affects the amount of time black students spend studying or how much they learn when they study. But we do have some indirect evidence.

Steele believes that “remedial” programs for black undergraduates can exacerbate stereotype threat and depress blacks' academic performance. Working with colleagues at the University of Michigan, he initiated an alternative program for a random sample of all students entering in the early 1990s. This program described itself as *more* demanding than the normal first-year curriculum. It seems to have boosted black students' undergraduate grades not just while they were in the program but afterwards.<sup>82</sup>

80. Frederick Vars and William Bowen's finding in chapter 13 that grade disparities between black and white undergraduates increase as their SAT scores increase is consistent with the hypothesis that stereotype threat is more of a problem among high-achieving blacks.

81. Osborne (forthcoming) presents evidence that the correlation between high school grades and self-esteem falls during the high school years for blacks but not whites.

82. Steele and others (1997).

This outcome suggests that the best way to improve black undergraduates' performance may be to treat them like everyone else and expect more than we do now. This may also be the best way to counter the perception that blacks are less intelligent than whites—a perception that has faded over the past twenty years but is still widespread.<sup>83</sup> Ferguson's review of research on teachers' expectations suggests that this logic also applies to younger black students.

### Test Scores and College Admissions

A number of American colleges limit admission to students who, by one standard or another, seem to be in the top 5 percent of their cohort. If these colleges based their admissions decisions entirely on test performance and only accepted students who scored above the 95th percentile, less than 2 percent of their students would be black.<sup>84</sup> In practice, most colleges use a combination of test scores and high school grades to predict applicants' college grades. This increases blacks' chances of qualifying for a selective college on the basis of academic merit alone, but if predicted college grades were the sole criterion for choosing among applicants, blacks would remain badly underrepresented in selective colleges.

83. The 1990 General Social Survey (GSS) asked respondents to rate various groups on a scale that ran from 1 ("unintelligent") to 7 ("intelligent"). Among white respondents, 31 percent gave blacks a score of 3 or less, 16 percent gave southern whites a score of 3 or less, and 6 percent gave all whites a score of 3 or less. We do not know how these ratings have changed over time, but we do have trend data on a related question. Since 1977 the GSS has repeatedly asked respondents why blacks have worse jobs, income, and housing than whites. One possible answer is that "most blacks have less in-born ability to learn." In 1977, 26 percent of whites agreed that this was one possible explanation for black poverty. By 1994, just before the publication of *The Bell Curve*, only 14 percent agreed. In 1996, only 10 percent agreed. This question almost certainly understates the percentage of whites who entertain doubts about blacks' intellectual ability. To begin with, Americans are far more likely to blame poverty on laziness than stupidity, and the GSS shows that this pattern holds for black poverty as well. This means that even whites who doubt blacks' intellectual ability may not think this is an important factor in explaining black poverty. In addition, some whites probably think blacks are unintelligent for environmental reasons, and others are probably reluctant to report politically incorrect views. Nonetheless, we believe the downward trend is real. We are indebted to Joseph Swingle for analyzing the GSS for us.

84. Roughly 15 percent of school-age children are black. Hedges and Nowell show that whites were ten times more likely than blacks to score in the top 5 percent of the national test score distribution. The estimate in the text assumes that the ratio would still be at least eight to one if we compared all nonblacks to blacks.

Fortunately, selective private colleges have never based their admissions decisions solely on predicted grades. They also favor applicants who are good at football, who edited their high school newspaper, who play the trombone, or who have extremely rich fathers. When selective colleges began trying to increase their black enrollment in the late 1960s, they simply added race to the list of attributes that could push an otherwise marginal applicant above the threshold for admission. This policy was always controversial, but it survived more or less unchanged until the mid-1990s. In 1995 the University of California's Board of Regents voted to ban racial preferences in admissions. In 1996 the Fifth Circuit Court of Appeals banned racial preferences throughout Texas, Louisiana, and Mississippi.<sup>85</sup> Such bans may well spread in the years ahead.

Thomas Kane analyzes the extent of racial preferences in chapter 12, using data from the 1980s. Most American colleges admit most of their applicants. These colleges do not need racial preferences to maintain a racially diverse student body. Kane finds that such preferences are important only at the most selective fifth of four-year colleges. An applicant's race makes a big difference at these institutions. In 1982 selective colleges were as likely to admit a black student with total SAT scores of 1100 and a B+ average as they were to admit a white student with total SAT scores of 1300 and an A- average.<sup>86</sup> Kane does not present comparable data for the 1990s, but we doubt that the extent of racial preferences changed much between 1982 and 1995. The situation has certainly changed since 1995 in California, Texas, Louisiana, and Mississippi. It may also have changed elsewhere.

Advocates of color-blind admissions usually argue that racial preferences are unfair to whites and Asians. To support this argument, they often point out that blacks who attend selective colleges earn lower undergraduate grades and are less likely to graduate than their white and Asian classmates. In some cases, advocates of color-blind admissions go further and argue that racial preferences are actually a disservice to their nominal beneficiaries, who would be better off at less demanding institutions. If this were true, racial preferences would certainly be hard to defend. But it seems to be false.

85. *Hopwood v. State of Texas*, 95 F 3d 53; U.S. App. Lexis 22891.

86. Estimated from table 12-2 of chapter 12 by Thomas Kane. The effect of race on an applicant's chance of admission obviously varies from one selective institution to the next, but Kane does not try to estimate the extent of this variation.

When Kane tracks high school students matched on SAT scores, high school grades, and family background, those who attend more selective colleges have about the same undergraduate grades as those who attend less selective colleges. If a B at Stanford or Northwestern represents a higher level of accomplishment than a B at San Jose State or Northern Illinois, Kane's findings imply that marginal students are learning more at selective colleges.

Attending a more selective college does not reduce a student's chance of graduating, either. Holding applicants' academic qualifications constant, Kane finds that those who attend more selective colleges are actually *more* likely to graduate. Perhaps students at selective colleges are more highly motivated to begin with. Perhaps they stick around because life at a selective college is more interesting or because a degree from Stanford or Northwestern is worth more than a degree from San Jose State or Northern Illinois. (Among students who looked similar at the end of high school, those who attended a more selective college earned a little more when they were in their late twenties, but this could be because they were more ambitious to begin with.)

Selective colleges often justify their exclusiveness on the grounds that unusually talented students need an unusually demanding curriculum, which would merely frustrate less talented students. This argument may well be correct, but it has never been tested empirically, because the United States has never collected data on how much students learn in different colleges. At the elementary school level, where we do have data, putting high- and low-ability children in separate classes does not seem to have much impact on how much they learn. But twenty-year-olds probably learn more from their classmates than ten-year-olds do, and studies of elementary school ability grouping measure mastery of very simple skills. Kane finds that the *economic* value of attending a selective college is actually higher for students with low SAT scores than for students with high SAT scores. This could be because selective colleges only take low-scoring applicants who have other virtues, but we doubt it.

Kane's findings suggest to us that when all else is equal students usually benefit from attending the most selective college that will admit them. Selective colleges cannot, by definition, admit everyone who would benefit from attending. If they did, they would no longer be selective, and the value of attending them would fall. At present, selective colleges allocate scarce places by trying to balance two ill-defined ideals: merit and diversity. Those who oppose racial preferences want selective colleges to allocate places on merit alone. For some, this means relying entirely on predicted college

grades. For others, it leaves room for football players, trombone players, and student newspaper editors.

The claim that selective colleges should only consider “merit” could mean that colleges should only consider an applicant’s past accomplishments, treating admission as a prize that goes to the most deserving. Most college admissions offices view merit somewhat differently. They think their task is to predict applicants’ future accomplishments. These two versions of “merit” have very different implications. A young man from a wealthy family is likely to exert a lot of influence on the world, even if his SAT scores are not especially impressive. He may also give the college a lot of money. Few selective colleges think they should ignore either of these facts. A college that sees itself as training America’s future leaders also knows that excluding applicants with SAT scores below 1300 will exclude almost all the nation’s future black leaders. Selective colleges are also reluctant to ignore this risk, since they do not want to position themselves as “white” (or even “white and Asian”) institutions in an increasingly multiracial society. Faced with a difficult problem that has no perfect solution, one might expect a society committed to *laissez-faire* to let every college pursue its own conception of virtue mixed with self-interest. But when racial preferences are the issue, Americans often lose their enthusiasm for *laissez-faire*.

Frederick Vars and William Bowen also explore the consequences of racial preferences in chapter 13, using data on all the students who entered eleven highly selective colleges and universities in 1989. Like Kane, they find that blacks earn lower college grades than whites with the same SAT scores, although the magnitude of the discrepancy varies from campus to campus. Black students’ low grades are not attributable to the fact that they come from families with less money or less schooling than the average white. Among applicants with the same SAT scores and high school grades, the measures of family background that college admissions offices collect do not predict college grades.<sup>87</sup>

Vars and Bowen’s most disturbing finding is that the black-white disparity in college grades is widest among students with high SAT scores. Elite colleges apparently have particular difficulty engaging their most promising black students.<sup>88</sup> This could be because black students tend to social-

87. Kane controls a more extensive list of SES measures in his analysis of the High School and Beyond data. He too finds that SES has a negligible effect on college grades once SAT scores and high school grades are controlled, although he does not display the coefficients in chapter 12.

88. Steele and others’ (1997) intervention program at the University of Michigan yielded its largest benefits for high-scoring black students.

ize with one another. A well-prepared black student may therefore have worse-prepared friends and study partners than an equally well-prepared white student. If racially delimited friendship patterns depress promising black students' grades, the problem might diminish if selective colleges set a higher threshold for admitting black applicants. But color-blind admissions could also have the perverse effect of forcing selective colleges to admit only those minority students whom they currently have the greatest difficulty motivating.

Faced with such uncertainty, we need more information. Over the past generation, some colleges have kept the black-white test score gap quite small while others have allowed it to grow very large. According to *The Bell Curve* the SAT gap at selective colleges in 1990–92 ranged from a high of 288 points at Berkeley to a low of 95 points at Harvard.<sup>89</sup> If large racial preferences have any effect on black students, positive or negative, we should have been able to measure this effect by comparing places like Berkeley to places like Harvard. Because colleges do not normally report the extent to which they favor black applicants, we have no studies of this kind.<sup>90</sup>

### Test Scores and Wages

Many American whites believe that large employers favor blacks over whites when making hiring and promotion decisions because such employers want to “make their numbers look good.” Many American blacks, in contrast, assume that they have to be more qualified than any white applicant to get a job or promotion. Both groups are presumably right some of the time. Some employers clearly favor black applicants for some jobs. Others clearly favor whites for some jobs. Such evidence as we have suggests, however, that whites are favored considerably more often than blacks.<sup>91</sup> The obvious question is why.

William Johnson and Derek Neal investigate test scores' contribution to racial differences in wages and employment in chapter 14. Their findings vary dramatically by gender. When they compare all black and white women in their late twenties, black women work about as many hours a

89. Herrnstein and Murray (1994). Kane reproduces the relevant table in chapter 12.

90. The report from which Herrnstein and Murray (1994) took most of their estimates of the black-white SAT gap is only available to participating institutions' admissions offices, not to social scientists.

91. Turner, Fix, and Struyk (1991) show that when matched black and white “testers” apply for jobs listed in the “Help Wanted” section of the newspaper, whites are favored more often than blacks.



year as white women but earn about 17 percent less per hour.<sup>92</sup> When Johnson and Neal compare women who had similar AFQT scores in adolescence, this picture changes dramatically. Black women work 15 percent more hours and earn 5 percent more per hour than white women with the same scores.<sup>93</sup> These findings could conceivably mean that employers prefer black women to equally skilled white women, but few employers say that is the case.<sup>94</sup> A more plausible explanation is that black women's small apparent wage advantage is a product of either sampling error or racial differences in geographic location, and that black women work more hours than equally skilled white women because black men work less, earn lower wages, and are less likely to live with their children.<sup>95</sup>

Among men, the picture is quite different. Black men in their late twenties work 30 percent fewer hours than whites and earn 24 percent less per hour. As a result, black men earn about half as much per year as white men.<sup>96</sup> When Johnson and Neal compare men with the same adolescent AFQT scores, black men still work 20 percent fewer hours than whites and earn 9 percent less per hour.<sup>97</sup> This means that the skills measured by the AFQT explain only half the racial disparity in men's annual earnings. We do not know why black men work less than white men with comparable cognitive skills, but the fact that this pattern does not hold among women rules out many possible explanations. One plausible hypothesis is that employers have more negative stereotypes of black men than of black women. An-

92. The coefficients of race in equations predicting the log of mean hourly wages and mean annual earnings for 1990–92 are  $-0.277$  and  $-0.272$ , so the implied coefficient of race in an equation predicting these individuals' hours is  $.005$ . This estimate omits women who did not work at any time during 1990, 1991, or 1992, who constituted 20 percent of the total sample.

93. The apparent effect of test scores could reflect the fact that family background influences both test scores and wages, but this bias is quite small (Korenman and Winship, forthcoming).

94. Kirschenman and Neckerman (1991).

95. Johnson and Neal do not analyze racial differences in hours worked by married women, but we would expect to find that black wives worked more than white wives with equally well paid husbands. Black wives' marriages are even less stable than white wives' marriages, so black wives have even more reason than white wives to keep working so that they will be able to support themselves and their children in the event of a divorce.

96. Johnson and Neal's estimate covers men who had earnings in any of three years (1990, 1991, or 1992). Most other racial comparisons are limited to men with positive earnings in a single year. Focusing on a single year excludes more men with weak labor force attachment and raises the black-white earnings ratio.

97. Table 14-1 of chapter 14 shows that controlling test scores reduces the logged black-white hourly wage gap among men from  $0.277$  to  $0.098$ . Table 14-5 shows that the coefficient for annual earnings drops from  $.640$  to  $.329$ . Subtraction yields the implied coefficients for the natural log of annual hours, which are  $.363$  without test scores controlled and  $.231$  with test scores controlled.

other possibility is that black men are less willing than black women to take the low-wage jobs that America now offers unskilled workers of both sexes.

### Can We Explain More of the Gap?

The evidence summarized in the first eleven chapters of this book shows that traditional explanations for the black-white test score gap do not work very well. If genes play any role, it is probably quite small. Poverty probably plays some role, but it too is modest. School desegregation probably reduced the black-white gap in the South during the 1970s, and desegregating northern elementary schools might raise blacks' reading scores today, but the gain would be modest. Reducing class size in the early grades would probably raise scores more for blacks than whites, and reducing class size in later grades might help preserve these black gains, although this latter conclusion is based on conjecture rather than firm evidence. Screening teachers for verbal and mathematical competence is also likely to raise black children's scores.

The United States ought to be conducting large-scale experiments aimed at reducing uncertainty about the effects of schools' racial mix, class size, teacher selection systems, ability grouping, and many other policies. We do such experiments to determine the effects of different medical treatments, different job training programs, and many other social interventions. But the U.S. Department of Education, which should in principle be funding experiments from which every state and school district would benefit, has shown almost no interest in this approach to advancing knowledge about education. The most important piece of educational research in the past generation, the Tennessee class-size experiment, was funded by the Tennessee legislature, not the U.S. Department of Education. Experimental assessments of other educational policies that have a major impact on school spending—salary levels, teacher selection systems, education for the physically and mentally disabled, and bilingual education, for example—have been almost nonexistent.

If we did more experiments, we might eventually develop better theories. At present, theorizing about the causes of the black-white gap is largely a waste of time, because there is no way to resolve theoretical disagreements without data that all sides accept as valid. Most theories about human behavior start out as hunches, anecdotes, or ideological predispositions. Such theories improve only when they have to confront evidence that the theorist cannot control. In education, that seldom happens.

Our best guess is that successful new theories about the causes of the black-white gap will differ from traditional theories in at least three ways:

—Instead of looking at families' economic and educational resources, successful theories will probably pay more attention to the way family members and friends interact with one another and with the outside world. A good explanation of why white four-year-olds have bigger vocabularies than black four-year-olds is likely to focus on how much parents talk to their children, how they deal with their children's questions, and how they react when their children either learn or fail to learn something, not on how much money the parents have in the bank.<sup>98</sup>

—Instead of looking mainly for resource differences between predominantly black and predominantly white schools, successful theories will probably pay more attention to the way black and white children respond to the same classroom experiences, such as having a teacher of a different race or having a teacher with low expectations for students who read below grade level.

—Instead of trying to trace the black-white test score gap to economic or demographic influences, successful theories will probably have to pay more attention to psychological and cultural differences that are harder to describe accurately and therefore easy to exaggerate. Collecting accurate data on such differences would require a massive investment of effort, perhaps comparable to what psychologists invested in developing cognitive tests during the first half of the twentieth century. It would also require far closer cooperation between psychologists, ethnographers, and survey researchers than one ordinarily sees in academic life.

### Can We Narrow the Gap?

We have argued that reducing the black-white test score gap would do more to move America toward racial equality than any politically plausible alternative. This argument rests on two problematic premises: first, that we know how to reduce the gap, and second that the policies required to reduce the gap could command broad political support.

When readers try to assess what experts know, they should start by drawing a sharp distinction between policies that are expensive but easy to implement and policies that are cheap but hard to implement. (Most policies that are cheap, easy to implement, and clearly effective—like teaching

98. See, for example, Hart and Risley (1992).

children the alphabet or the multiplication tables—are already universal.) The two policies that seem to us most likely to combine effectiveness with ease of implementation are cutting class size and screening out teachers with weak academic skills. Cutting class size is clearly expensive, although we might be able to make some progress by moving existing teachers back into regular classes.<sup>99</sup> Selecting teachers with higher test scores would not be expensive in districts that now have more applicants than openings, but it would require higher salaries in many big cities.<sup>100</sup>

When educators look for less expensive ways of improving black children's achievement, they usually find themselves considering proposals that are quite difficult to implement. Raising teachers' expectations is not inherently expensive, for example, but how does a school administrator do it? Big-city school districts are besieged by advocates of curricular innovation who claim their programs will raise black children's test scores. These programs usually require complex and relatively subtle changes in classroom practice. School boards and administrators cannot impose such changes by decree, the way they can reduce class size or require new teachers to pass an exam. Nor can teachers make such changes by a single act of will, the way they might adopt a new textbook. As a result, schools seldom implement these programs in exactly the way their designers expected or intended. A program may work well initially, when it is closely supervised by a dedicated innovator, but may have no detectable effect when the innovator tries to bottle it and sell it off the shelf.<sup>101</sup>

Proposals for reducing the black-white test score gap also arouse passionate white resistance if they seem likely to lower white children's achievement. Both school desegregation and the elimination of academically

99. Taken at face value, Lewitt and Baker's (1997) data imply that during the hours when students were in class in 1991 about  $17/24 = 70$  percent of teachers were in class. In 1961 the figure was  $26/29 = 90$  percent. Part of this apparent reduction in teachers' classroom time may be spurious, however. Schools now have more specialized teachers for disabled children, bilingual instruction, and slow readers. These teachers' classes tend to be smaller than regular classes and may not be included in the estimates of average class size.

100. If wages did not rise in central cities, emphasizing teachers' test scores might actually widen the achievement gap between minority students in the central city and predominantly white students in the surrounding suburbs, since it might improve the quality of instruction more in districts that now have a surplus of applicants for every opening.

101. The program with the best-documented success record seems to be Robert Slavin's Success for All. Ferguson discusses this program in more detail in chapter 9. Success for All is more expensive than many of its competitors. For an excellent account of an apparently successful intervention that cost considerably less see Farkas (1996).

selective classes often arouse such fears. School desegregation probably raises black elementary school students' reading scores. Whether it lowers white scores is unclear. But once black enrollment in a neighborhood school expands past something like 20 percent, most white parents become reluctant to move into the neighborhood. If black enrollment remains low enough to keep whites comfortable, many blacks feel uncomfortable. There is no simple way out of this dilemma. It takes years of patient work to create stable, racially mixed schools with high black enrollments, and such schools remain unusual.

Ferguson's review suggests that the struggle over ability grouping at the elementary level is largely symbolic. Eliminating such classes would not do black children much good, and it would probably not do whites much harm either. At the secondary level, eliminating demanding courses seems ridiculous. We should be trying to get more black students to take such classes, not trying to eliminate them as an option for whites, who will respond by sending their children elsewhere. Any politically workable educational strategy for reducing the black-white test score gap has to promise some benefits for whites as well as blacks. Reducing class size, requiring greater academic competence among teachers, and raising teachers' expectations for students who have trouble with schoolwork all meet this test.

Although we believe that improving the nation's schools could reduce the black-white test score gap, we do not believe that schools alone can eliminate it. To begin with, competition for educational resources is fierce. The typical American voter might accept a system of educational finance that gave schools with unusually disadvantaged students 10 or 20 percent more money per pupil than the average school. But few Americans would accept a system that gave disadvantaged schools 50 or 100 percent more money than the average school. If smaller classes in disadvantaged schools improve children's reading skills, for example, more affluent parents will want smaller classes too. In a system of school finance that relies heavily on local funding, affluent parents who want smaller classes will usually be able to get them. Even ensuring equal funding for black and white schools is a constant struggle. Creating a system in which black schools get far more money than white schools is politically inconceivable.

Even if resources were not a constraint, the cognitive disparities between black and white preschool children are currently so large that it is hard to imagine how schools alone could eliminate them. Figure 1-1 showed that among three- and four-year-olds the typical black child's vocabulary

score falls below the 20th percentile of the national distribution. Relying entirely on educational reform to move such a child up to the 50th percentile does not strike us as realistic. If we want equal outcomes among twelfth-graders, we will also have to narrow the skill gap between black and white children before they enter school.

Broadly speaking, there are two ways of improving three- and four-year-olds' cognitive skills: we can change their preschool experiences and we can change their home experiences. Changing preschools is less important but easier than changing homes. Black preschoolers are concentrated in Head Start. At present, the program does not concentrate on teaching cognitive skills. Few Head Start teachers are trained to do this, and some oppose the idea on principle. Steven Barnett's review of research on preschool effects strongly suggests that cognitively oriented preschool programs can improve black children's achievement scores, even though the benefits fade somewhat as children age.<sup>102</sup> Getting Head Start to emphasize cognitive development should therefore be a higher priority than merely expanding its enrollment.

Parenting practices almost certainly have more impact on children's cognitive development than preschool practices. Indeed, changing the way parents deal with their children may be the single most important thing we can do to improve children's cognitive skills. But getting parents to change their habits is even harder than getting teachers to change. Like teachers, parents are usually suspicious of unsolicited suggestions. This is doubly true when the suggestions require fundamental changes in a parent's behavior. But once parents become convinced that changing their behavior will really help their children, many try to change, even when this is quite difficult. As a practical matter, whites cannot tell black parents to change their practices without provoking charges of ethnocentrism, racism, and much else. But black parents are not the only ones who need help. We should be promoting better parenting practices for *all* parents, using every tool at our disposal, from Head Start outreach programs and home visits by nurses to television sitcoms or anything else that might affect parents' behavior.<sup>103</sup>

A successful strategy for raising black children's test scores must also try to convince both blacks and whites that the gap is not genetic in origin.

102. Barnett (1995). Currie and Thomas (1995) find that Head Start's effect on children's test scores fades after a few years among blacks but not among whites. Currie and Thomas (1998) argue that this is because white children go on to better schools.

103. See Olds and others (1998) on the uneven cognitive effects of home visits by nurses.

This is not a simple task. Genetic variation does explain a substantial fraction of the variation in cognitive skills among people of the same race. But so does environmental variation. Once hereditarianism percolates into popular culture, it can easily become an excuse for treating academic failure as an inescapable fact of nature. Teaching children skills that do not seem to “come naturally” is hard work. If our culture allows us to avoid such work by saying that a child simply lacks the required aptitude to master the skill, both teachers and parents will sometimes jump at this as an excuse for not trying. This often makes everyone’s life pleasanter in the short run, but in the long run it is a formula for failure. Emphasizing heredity is likely to have especially negative consequences for African Americans, who start out behind whites and therefore need to work even harder than whites if they are to catch up.

The agenda we have sketched would not be easy to implement. We are not optimistic about expanding federal support for efforts of this kind. Popular distrust of federal education programs is now quite pervasive and shows no sign of receding. We are more optimistic about state and local efforts to narrow the black-white test score gap. Everyone recognizes that racial conflict is one of the nation’s most pressing and persistent problems. Other strategies for dealing with this problem, which emphasize the use of racial preferences to overcome the adverse effects of discrimination, the test score gap, or both, are clearly in political trouble. Public support for efforts to narrow the test score gap, while tempered by suspicion that “nothing works,” still seems fairly widespread. One reason is that the beneficiaries appear so deserving. Hardly anyone blames black first-graders’ limited vocabulary on defects of character or lack of ambition. First-graders of every race seem eager to please. That was why Lyndon Johnson placed so much emphasis on helping children in his original War on Poverty.

We recognize that few readers will find our sketchy agenda for reducing the black-white test score gap entirely persuasive. Such skepticism is completely reasonable. While we are convinced that reducing the gap is both necessary and possible, we do not have a detailed blueprint for achieving this goal and neither does anyone else. This book tries to update the reader’s knowledge about many aspects of the problem, but almost every chapter raises as many questions as it answers. This is partly because psychologists, sociologists, and educational researchers have devoted far less attention to the test score gap over the past quarter century than its political and social consequences warranted. Most social scientists have chosen safer topics and hoped the problem would go away. It didn’t. We can do better.

## References

- Ainsworth-Darnell, James, and Douglas Downey. Forthcoming. "Assessing the Oppositional Explanation for Racial/Ethnic Differences in School Performance." *American Sociological Review*.
- Armor, David. 1992. "Why Is Black Educational Achievement Rising?" *Public Interest*, no. 108 (Summer): 65–80.
- Barnett, Steven. 1995. "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes." *Future of Children* 5 (3): 25–50.
- Boozer, Michael, Alan Krueger, and Shari Wolkon. 1992. "Race and School Quality since *Brown v. Board of Education*." *Brookings Papers on Economic Activity (Microeconomics)*: 269–326.
- Bouchard, Thomas, and others. 1990. "Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart." *Science*, 250 (October 12): 223–28.
- Braungart, Julia, David Fulker, and Robert Plomin. 1992. "Genetic Mediation of the Home Environment during Infancy: A Sibling Adoption Study of the Home." *Developmental Psychology*, 28: 1048–55.
- Brody, Nathan. 1992. *Intelligence*. San Diego: Academic Press.
- Cherny, Stacey, and Lon Cardon. 1994. "General Cognitive Ability," in John DeFries, Robert Plomin, and David Fulker, eds., *Nature and Nurture in Middle Childhood*. Oxford: Blackwell.
- Chipuer, Heather, Michael Rovine, and Robert Plomin. 1990. "LISREL Modeling: Genetic and Environmental Influences on IQ Revisited." *Intelligence* 14 (1): 11–29.
- Coleman, James, and others. 1966. *Equality of Educational Opportunity*. Department of Health, Education, and Welfare.
- Currie, Janet, and Duncan Thomas. 1995. "Does Headstart Make a Difference?" *American Economic Review*, 85 (3): 341–64.
- . 1998. "School Quality and the Longer-Term Effects of Headstart," Working Paper 6362. Cambridge, Mass.: National Bureau of Economic Research.
- Cutright, Phillips. 1972. "Achievement, Mobility, and the Draft: Their Impact on the Earnings of Men," DHEW (SSA) 73-11854. Department of Health, Education, and Welfare.
- . 1974. "Academic Achievement, Schooling, and the Earnings of White and Black Men." *Journal of Social and Behavioral Sciences*, 20 (3): 1–18.
- Entwisle, Doris, and Karl Alexander. 1994. "Winter Setback: The Racial Composition of Schools and Learning to Read." *American Sociological Review*, 59 (June): 446–60.
- Entwisle, Doris, Karl Alexander, and Linda Steffel Olson. 1997. *Children, Schools, and Inequality*. Boulder: Westview Press.
- Evans, William N., Sheila Murray, and Robert Schwab. 1997. "School Houses, Court Houses, and State Houses after Serrano." *Journal of Policy Analysis and Management*, 16 (January): 10–31.
- Farkas, George. 1996. *Human Capital or Cultural Capital? Ethnicity and Poverty Groups in an Urban School District*. Aldine de Gruyter.
- Flynn, James. 1980. *Race, IQ, and Jensen*. London: Routledge and Kegan Paul.
- . 1984. "The Mean IQ of Americans: Massive Gains 1932 to 1978." *Psychological Bulletin*, 95 (1): 29–51.



- . 1987. "Massive IQ Gains in 14 Nations: What IQ Tests Really Measure." *Psychological Bulletin*, 101 (2): 171–91.
- Fordham, Signithia, and John Ogbu. 1986. "Black Students' School Success: Coping with the 'Burden of Acting White.'" *Urban Review*, 18(3): 176–206.
- Glass, Gene, and others. 1982. *School Class Size*. Beverly Hills: Sage Publications.
- Hanushek, Eric. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher*, 18 (4): 45–51.
- . 1998. "The Evidence on Class Size," occasional paper 98-1. University of Rochester, W. Allen Wallis Institute of Political Economy.
- Hartigan, John, and Alexandra Wigdor, eds. 1989. *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. Washington: National Academy Press.
- Hauser, Robert, and David Featherman. 1976. "Equality of Schooling: Trends and Prospects." *Sociology of Education*, 49: 99–120.
- Herrnstein, Richard, and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.
- Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling*. Academic Press.
- . 1987. "Schooling and Cognitive Development: Is There a Season for Learning?" *Child Development*, 58: 1151–60.
- Hopkins, Kenneth, and Glenn Bracht. 1975. "Ten-Year Stability of Verbal and Nonverbal IQ Scores." *Educational Research Journal*, 12: 469–77.
- Howard, Jeff, and Ray Hammond. 1985. "Rumors of Inferiority." *New Republic*, September 9: 18–23.
- Jencks, Christopher, and others. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. Basic Books.
- Jensen, Arthur. 1969. "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review*, 39: 1–123.
- Kirschenman, Joleen, and Kahryn Neckerman. 1991. "We'd Love to Hire Them But . . . : The Meaning of Race for Employers." In Christopher Jencks and Paul Peterson, eds. *The Urban Underclass*, pp. 203–32. Brookings.
- Korenman, Sanders, and Christopher Winship. Forthcoming. "A Reanalysis of *The Bell Curve*: Intelligence, Family Background, and Schooling." In Steven Durlauf and Samuel Bowles, eds., *Meritocracy and Society*. Princeton University Press.
- Lewitt, Eugene, and Linda Schumann Baker. 1997. "Class Size." *Future of Children*, 7 (3): 112–21.
- Loehlin, John. 1980. "Recent Adoption Studies of IQ." *Human Genetics*, 55: 297–302.
- Loehlin, John, Steven Vandenberg, and R. T. Osborne. 1973. "Blood Group Genes and Negro-White Ability Differences." *Behavior Genetics*, 3: 263–70.
- Loehlin, John, Gardner Lindzey, and J. N. Spuhler. 1975. *Race Differences in Intelligence*. San Francisco: W. H. Freeman.
- Loehlin, John, Joseph Horn, and Lee Willerman. 1989. "Modeling IQ Changes: Evidence from the Texas Adoption Project." *Child Development*, 60: 993–1004.
- Mayer, Susan. 1997. *What Money Can't Buy: Family Income and Children's Life Chances*. Harvard University Press.
- McCartney, Kathleen, Monica Harris, and Frank Bernieri. 1990. "Growing Up and Growing Apart: A Developmental Meta-Analysis of Twin Studies." *Psychological Bulletin*, 107 (2): 226–37.

- McLanahan, Sara, and Gary Sandefur. 1994. *Growing Up with a Single Parent*. Harvard University Press.
- Miller, L. Scott. 1995. *An American Imperative: Accelerating Minority Educational Advancement*. Yale University Press.
- Moore, Elsie. 1986. "Family Socialization and IQ Test Performance of Traditionally and Transracially Adopted Black Children." *Developmental Psychology*, 22 (3): 317–26.
- Mosteller, Frederick. 1995. "The Tennessee Study of Class Size in the Early Grades." *Future of Children*, 5 (2): 113–27.
- Mosteller, Frederick, and Daniel P. Moynihan, eds. 1972. *On Equality of Educational Opportunity*. Random House.
- Mosteller, Frederick, Richard Light, and Jason Sacks. 1996. "Sustained Inquiry in Education: Lessons from Skill Grouping and Class Size." *Harvard Educational Review*, 66 (4): 797–842.
- Mulkey, Lynn, Robert Crain, and Alexander Harrington. 1992. "One Parent Households and Achievement: Economic and Behavioral Explanations of a Small Effect." *Sociology of Education*, 65 (January): 48–65.
- Neisser, Ulrich, ed. 1998. *The Rising Curve: Long-Term Gains in IQ and Related Measures*. Washington: American Psychological Association.
- Ogbu, John. 1978. *Minority Education and Caste: The American System in Cross-Cultural Perspective*. Academic Press.
- . 1986. "The Consequences of the American Caste System." In Ulrich Neisser, ed., *The School Achievement of Minority Children: New Perspectives*, pp. 19–56. Hillsdale, N.J.: Erlbaum.
- Olds, David L., and others. 1998. "Prenatal and Infancy Home Visitation by Nurses: A Program of Research." Washington: American Enterprise Institute.
- Osborne, Jason. Forthcoming. "Race and Academic Disidentification." *Journal of Educational Psychology*.
- Phillips, Meredith. 1997. "Does School Segregation Explain Why African Americans and Latinos Score Lower than Whites on Academic Achievement Tests?" Paper prepared for the annual meeting of the American Sociological Association.
- Plomin, Robert. 1994. *Genetics and Experience: The Interplay between Nature and Nurture*. Newbury Park, Calif.: Sage.
- and John DeFries. 1980. "Genetics and Intelligence: Recent Data." *Intelligence* 4: 15–24.
- Plomin, Robert, and C. S. Bergeman. 1991. "The Nature of Nurture: Genetic Influence on 'Environmental' Measures." *Behavioral and Brain Sciences*, 14 (3): 373–427.
- Plomin, Robert, and Stephen Petrill. 1997. "Genetics and Intelligence: What's New?" *Intelligence*, 24 (1): 53–77.
- Plomin, Robert, John DeFries, and David Fulker. 1988. *Nature and Nurture in Infancy and Early Childhood*. Cambridge University Press.
- Rao, D. C., and others. 1982. "Path Analysis under Generalized Assortative Mating." *Genetical Research*, 39: 187–98.
- Scarr, Sandra, and Richard Weinberg. 1976. "IQ Test Performance of Black Children Adopted by White Families." *American Psychologist*, 31: 726–39.
- . 1978. "The Influence of 'Family Background' on Intellectual Attainment." *American Sociological Review*, 43 (October): 674–92.

- Scarr, Sandra, and others. 1977. "Absence of a Relationship between Degree of White Ancestry and Intellectual Skills within a Black Population." *Human Genetics*, 39: 69–86.
- Snyderman, Mark, and Stanley Rothman. 1987. "Survey of Expert Opinion on Intelligence and Aptitude Testing." *American Psychologist*, 42 (2): 137–44.
- . 1990. *The IQ Controversy, the Media, and Public Policy*. New Brunswick, N.J.: Transaction Books.
- Steele, Claude. 1992. "Race and the Schooling of Black Americans." *Atlantic Monthly*, April: 68–78.
- . 1997. "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist*, 52 (6): 613–29.
- Steele, Claude, and others. 1997. "African American College Achievement: A 'Wise' Intervention." Stanford University Department of Psychology.
- Tuddenham, Reed. 1948. "Soldier Intelligence in World Wars I and II." *American Psychologist*, 3:54–56.
- Turner, Margery, Michael Fix, and Raymond Struyk. 1991. *Opportunities Denied, Opportunities Diminished*. Washington: Urban Institute.
- Vernon, Philip. 1979. *Intelligence: Heredity and Environment*. San Francisco: W. H. Freeman.
- Weinberg, Richard, Sandra Scarr, and Irwin Waldman. 1992. "The Minnesota Transracial Adoption Study: A Follow-Up of IQ Performance at Adolescence." *Intelligence*, 16: 117–35.
- Wigdor, Alexandra, and Bert Green, eds. 1991. *Performance Assessment in the Workplace*, vol. 1. Washington: National Academy Press.
- Willerman, Lee, Alfred Naylor, and Ntinos Myrianthopoulos. 1974. "Intellectual Development of Children from Interracial Matings: Performance in Infancy and at 4 Years." *Behavior Genetics*, 4 (1): 83–90.
- Winship, Christopher, and Sanders Korenman. 1997. "Does Staying in School Make You Smarter? The Effect of Education on IQ in *The Bell Curve*." In Bernie Devlin and others, eds., *Intelligence, Genes, and Success*, pp. 215–34. Copernicus.