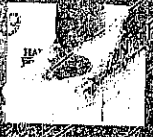
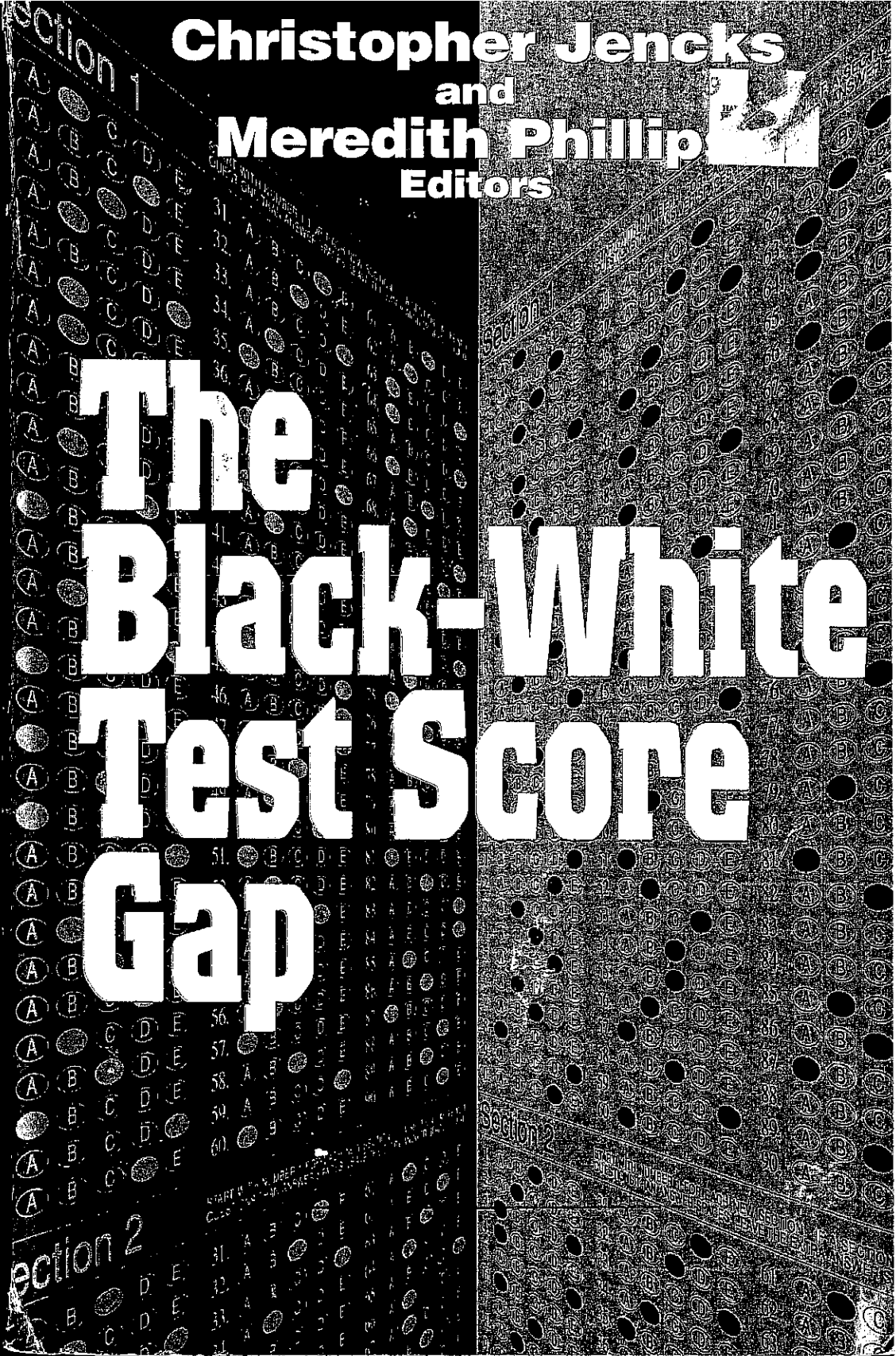


Christopher Jencks
and
Meredith Phillip
Editors



The Black-White Test Score Gap



1997. "Enhancing the Validity and Usefulness of the NELS:88 Mathematics Achievement to 12th Grade." *Journal* 34: 124-50.

Spuhler. 1975. *Race Differences in Intelligence*.

1995. *Psychometric Report for the NELS:88*. Washington: Office of Educational Research and Statistics.

1997. *Department of Education*.

1997. "The National Education Longitudinal Study: Academic Achievement."

1997. "The National Education Longitudinal Study: Academic Achievement."

1994. "Combining Estimates of Effect Size." *Research Synthesis*, pp. 261-81.

RONALD F. FERGUSON

8

Teachers' Perceptions and Expectations and the Black-White Test Score Gap

AFRICAN-AMERICAN CHILDREN arrive at kindergarten with fewer reading skills than whites, even when their parents have equal years of schooling. In an ideal world, schools would reduce these disparities. Unfortunately, national data show that, at best, the black-white test score gap is roughly constant (in standard deviations) from the primary through the secondary grades.¹ At worst, the gap widens.² Among blacks and whites with equal current scores, blacks tend to make less future progress. This is the first of two chapters on how schools might affect this story. It examines evidence for the proposition that teach-

Thanks to Karl Alexander, William Dickens, James Flynn, Christopher Jencks, Meredith Phillips, and Jason Snipes for helpful discussions and comments on earlier drafts. I am also grateful to Lee Jussim and Meredith Phillips for calculations that they conducted at my request for this chapter. Jason Snipes provided able research assistance.

1. See chapter 7. The black-white gap in skills at the beginning of primary school is smaller for more recent cohorts.

2. Existing evidence on group-level disparity at the mean across grade levels within a cohort is not entirely clear, because of measurement issues and data problems. See chapter 7 above for a discussion of methodological decisions that determine whether the gap appears to be constant or to grow wider over time within a cohort.

ers' perceptions, expectations, and behaviors interact with students' beliefs, behaviors, and work habits in ways that help to perpetuate the black-white test score gap.³

No matter what material resources are available, no matter what strategies school districts use to allocate children to schools, and no matter how children are grouped for instruction, schoolchildren spend their days in social interaction with teachers and other students. As students and teachers immerse themselves in the routines of schooling, perceptions and expectations both reflect and determine the goals that they set for achievement; the strategies they use to pursue the goals; the skills, energy, and other resources they use to implement the strategies; and the rewards they expect from making the effort. These should affect standardized scores as well as other measures of achievement.

It is a common, if controversial, assumption that teachers' perceptions, expectations, and behaviors are biased by racial stereotypes. The literature is full of seeming contradictions. For example, Sara Lawrence Lightfoot writes: "Teachers, like all of us, use the dimensions of class, race, sex, ethnicity to bring order to their perception of the classroom environment. Rather than teachers gaining more in-depth and holistic understanding of the child, with the passage of time teachers' perceptions become increasingly stereotyped and children become hardened caricatures of an initially discriminatory vision."⁴

Similarly, Reuben Baron, David Tom, and Harris Cooper argue that "the race or class of a particular student may cue the teacher to apply the generalized expectations, therefore making it difficult for the teacher to develop *specific* expectations tailored to individual students. In this manner, the race or class distinction among students is perpetuated. The familiar operation of stereotypes takes place in that it becomes difficult for minority or disadvantaged students to distinguish themselves from the generalized expectation."⁵

3. Whenever possible, I present results in terms of "effect sizes" measured in standard deviation units. For example, if one group of students experiences a particular treatment and an other's equivalent control or comparison group does not, the effect size of the treatment on test scores is the difference between average scores for the two groups after the treatment, divided by the pooled standard deviation of scores. For an outcome that is normally distributed, an effect size of 0.20 moves a student from the fiftieth to the fifty-eighth percentile; an effect size of 0.50 moves the student to the sixty-ninth percentile; and an effect size of 0.80 moves the student to the seventy-ninth percentile.

4. Lightfoot (1978, pp. 85-86).

5. Baron, Tom, and Cooper (1985, p. 251).

On the other side, Jerome Brophy doubts that bias is important: "Few teachers can sustain grossly inaccurate expectations for many of their students in the face of daily feedback that contradicts those expectations."⁶ Emil Haller adds: "Undoubtedly there are some racially biased people who are teachers. . . . However . . . the problem does not seem to be of that nature. Conceiving it so is to confuse the issue, to do a serious injustice to the vast majority of teachers, and ultimately to visit an even more serious one on minority pupils. After all . . . children's reading skills are not much improved by subtly (and not so subtly) labeling their teachers racists."⁷

Some aspects of this debate are substantive, but others are semantic. The chapter begins by distinguishing among alternative definitions of racial bias and reviewing evidence on teachers' perceptions and expectations. Later sections address the ways in which teachers' and students' behaviors might be both causes and consequences of racially disparate perceptions and expectations regarding achievement, and might therefore contribute to perpetuating the black-white test score gap.

Bias in Teachers' Perceptions and Expectations

Expectations, perceptions, and behaviors that look biased under one criterion often look unbiased under another. However, researchers who study racial bias seldom evaluate their findings by more than a single standard. The discourse that results can be quite perplexing. One body of literature alleges bias and another denies it, but much of this disagreement is really over what is meant by "bias." At least three different conceptions of bias appear in this debate.

Bias is deviation from some benchmark that defines neutrality, or lack of bias. One type of benchmark is "unconditional" race neutrality. By this criterion, teachers who are unbiased expect the same, on average, of black and white students. A second type of benchmark is "conditional" race neutrality—that is, conditioned on observable, measurable criteria. For example, unbiased teachers should expect the same of black and white students on the condition that they have the same past grades and test scores. The third type of benchmark is conditioned not on past performance but on

⁶ Brophy (1985, p. 304).

⁷ Haller (1985, p. 481) is commenting on racial disparity in ability group assignments. See chapter 2 below for more discussion of ability grouping, including Haller's findings.

unobserved potential. It requires neutrality—for example, equal expectations and aspirations—in regard to blacks and whites who have equal potential. Insofar as “potential” differs from past performance, however, it is difficult to prove. Assuming that black and white children are born with the same potential (which seems a fair assumption), there is no distinction at birth between unconditional race neutrality and neutrality conditioned on unobserved potential. However, disparities in potential may develop as children grow older; recent literature on brain development, for example, suggests that experience alters potential. Therefore unconditional race neutrality may or may not remain the best approximation to neutrality conditioned on unobserved potential.

Unconditional Race Neutrality

Unconditional race neutrality requires that teachers’ perceptions, expectations, and behaviors be uncorrelated with students’ race. By this definition, an unbiased perception, expectation, or treatment has the same average value for all racial groups. This benchmark for racial bias is the standard in experimental studies. Such studies typically find that teachers are racially biased.

In experimental studies researchers systematically manipulate information about students. Although race is included, the sample is selected to avoid any correlation between race and the other characteristics noted.⁸ In a typical experiment, teachers receive information about students in written descriptions, photographs, videotapes—or occasionally, real children, who act as the experimenter’s confederates. The teachers then predict one or more measures of ability or academic performance for each student. If the experiment is run well, the teachers do not discern that race is a variable in the experiment or that the real purpose is to assess their racial biases.

Baron, Tom, and Cooper conduct a meta-analysis of experimental studies that focus on teachers’ expectations, sixteen of which deal with race. Teachers have higher expectations for white students in nine of the studies.

8. This immediate discussion concerns experiments. However, this first type of benchmark is also used in naturalistic settings. Specifically, in the absence of reliable information about individual ability, which to base a benchmark that is not unconditionally race neutral, unconditional racial neutrality may seem the only morally defensible alternative.

9. Baron, Tom, and Cooper (1985).

and for blacks in one study. The differences are statistically significant in five of these studies, all favoring whites. The remaining six studies do not report which group is favored. In these studies, the differences are statistically insignificant. Overall, the hypothesis of identical expectations for black and white students is clearly rejected ($p < 0.002$).¹⁰

An interesting study by Debra DeMeis and Ralph Turner is not included in Baron, Tom, and Cooper’s meta-analysis but supports their conclusion.¹¹ Sixty-eight white female elementary school teachers, with an average of seven years of teaching experience, were selected from summer school classes at a university in Kentucky during the 1970s. They were played tapes of male fifth-graders responding to the question, “What happened on your favorite TV show the last time you watched it?” and each and Turner asked teachers to rate the taped responses for personality, quality of response, current academic abilities, and future academic abilities. The race of the student in the picture was a statistically significant predictor for each of the four outcomes ($p < 0.0001$).¹²

If the benchmark is unconditional race neutrality, teachers are found to hold racially biased expectations. What should one make of this pervasive racial bias? Consider people who learn from real life that when one flips a coin the odds of getting heads are 60:40. Place these people in an experimental situation where, unknown to them, the odds have been set to 50:50. If each person is then given only one toss of the coin, will their predictions be unbiased? In an environment where real differences in per-

10. Baron, Tom, and Cooper (1985) report that effect sizes could be retrieved for only six of the sixteen studies. In these six studies the black-white differences in teacher expectations average half a standard deviation. If nine of the other studies are assumed to have effect sizes of zero and the one with a significant result but no effect size is assumed to have an effect size of 0.36, then the average effect size across all sixteen studies is 0.22.

11. DeMeis and Turner (1978).

12. To compute the effect sizes for each outcome, I use the standard deviation among blacks as the denominator, since the pooled standard deviation is not reported. The standard deviation among whites is virtually the same as that among blacks. Among students speaking standard English, effect sizes for the black-white differences in personality, quality of response, current academic abilities, and English, the analogous numbers are 0.57, 0.52, 0.55, and 0.44 standard deviations, respectively. For black English, the analogous numbers are 0.34, 0.44, 0.23 and 0.14 standard deviations. The fact that effect sizes are smaller for tapes on which students spoke black English is not surprising, since speaking black English would be an especially negative signal for a white student. Across the four outcomes, white-race effect sizes for black English versus standard English range from 0.23 to 0.45 for blacks and from 0.55 to 0.74 for whites, in all cases favoring standard English. All effect sizes reported in this footnote are calculated from numbers in DeMeis and Turner (1978, table 2).

formance between blacks and whites are the norm, if the benchmark for declaring expectations unbiased is unconditional race neutrality, biased expectations are what one should expect.¹³ For the same reasons, this type of bias is also pervasive in naturalistic studies—that is, studies in real classrooms, without experimental controls.

Experimental research of this kind establishes that teachers believe certain stereotypes and use the stereotypes in one-time encounters with experimental targets. But it does not establish that the stereotypes would represent biased estimates of the average if they were applied in real classrooms outside the experimental setting. Nor does it prove that teachers in real classrooms treat students inappropriately, or that their stereotypes prevent them from forming accurate perceptions about individual students.

Evidence on Accuracy

For at least two decades, scholars in education have emphasized that teachers' contemporaneous perceptions of students' performance, as well as their expectations of students' future performance, are generally accurate.¹⁴ For example, it has been found that first grade teachers can learn enough about children in the first few weeks of school to predict with some accuracy their rank order on examinations held at the beginning of second grade.¹⁵ Once set, teachers' expectations do not change a great deal. This may be because their early impressions of proficiency are accurate, and the actual rank order does not change much.

There are several possible reasons for stability in rank ordering. First, teachers' perceptions and expectations might be relatively inflexible. Self-fulfilling expectation effects, discussed below, will typically be strongest for the teachers whose expectations are least flexible.¹⁶ For these teachers, cor-

13. For whatever reasons, average scores on standardized examinations tend to be lower for blacks than for whites. Hence in general the most accurate prediction is that whites will have higher scores.

14. See, for example, Good (1987); Egan and Archer (1985); Mirman (1985); Hoge and Butcher (1988); Monk (1983); Redulla, Ahrasian, and Madans (1980).

15. See Brophy and Good (1974, table 6.1), which presents correlations from Evertson, Brophy, and Good (1972).

16. Ecles and Whigfield (1985), for example, provide a line of reasoning that might support this outcome. Essentially, in order to deviate from a previously established trajectory, the student may need support from the teacher in beginning the process of change. If the teacher continues to treat the student as he or she did before the change, the student may decide that the environment is not sufficiently responsive to attain the new goal, feel a lack of control, and return to the old ways.

relations between beginning-of-year and end-of-year assessments should be among the highest.¹⁷ A second reason for stability might be that few students try hard to change their positions. A third might be that the pace and style of standard teaching offer few effective opportunities for students who are behind to catch up.¹⁸ Most evidence about the accuracy of teachers' perceptions comes from correlations between teachers' predictions and actual test scores, which typically range between 0.50 and 0.90.¹⁹ At least at the low end of this range, one might alternatively focus on the inaccuracy of the predictions, in "glass half empty" fashion.

I know of only three studies that report separate correlations for blacks and whites. Haller finds that the correlation between teachers' subjective assessments of the reading proficiency of fourth, fifth, and sixth graders and students' scores on the Comprehensive Test of Basic Skills is 0.73 for whites and 0.74 for blacks.²⁰ Jacqueline Irvine asked teachers to rank 213 fifth, sixth, and seventh graders on general academic ability during the second, tenth, and final weeks of the 1983-84 school year. Correlations between these ratings and scores on the California Achievement Test are similar for blacks and whites.²¹ Similarly, Margie Gaines finds that teachers' predictions of performance on the Iowa Test of Basic Skills are as accurate for black students as for whites.²²

17. Experimental studies that expose teachers to different sequences of factors show that their expectations are sufficiently flexible to remain accurate as new information becomes available; see, for example, Shavelson, Cadwell, and Izu (1977). The pattern of flexibility among teachers in real classrooms is not known.

18. Indeed, Guskey (1982) finds that improvements in teacher responsiveness reduce the accuracy of teachers' early predictions for end-of-semester performance; see discussion in text below.

19. See Egan and Archer (1985); Irvine (1985); Brophy and Good (1974); Evertson, Brophy, and Good (1972); Willis (1972).

20. Haller (1985, note 4). This study covers forty-nine teachers and 934 fourth, fifth, and sixth graders in five cities across four census regions.

21. See Irvine (1990, p. 77), discussing findings presented first in Irvine (1985). The correlations between second-week rankings and end-of-year test scores are 0.63 for white males and 0.62 for black males. The correlation for black males dipped in the tenth week, but had returned to the same range as for whites by the end of the school year. Irvine emphasizes this difference in the pattern for black and white boys. It seems to me, however, that similarity is the more salient finding: of three comparisons for boys, in only one (boys at the tenth week) is the racial difference notable, and there is no significant racial difference in the three comparisons for girls. Some teachers in Irvine's study were consistently more accurate than others. For the least accurate teacher, the correlations moved from 0.11 for the second week to 0.56 for the end of the year. At the high end, one teacher had correlations of 0.91, 0.92, and 0.89 for the second week, tenth week, and end of the year, respectively.

22. Gaines (1990).

The similarity in correlations for blacks and whites means that the rank order of achievement among blacks is as stable as that among whites, and that teachers achieve similar accuracy in assessing both racial groups. It does not, however, imply that teachers' perceptions or expectations have the same impact on blacks and whites.²³ Neither does it mean that teachers are racially unbiased. In this context, accuracy is not always the opposite of bias. If self-fulfilling prophecy were always perfect, for example, each student's performance would be exactly what the teacher expected. Therefore if expectations were biased, outcomes would be perfectly predicted but biased.

Race Neutrality Conditioned on Observables

The second type of benchmark for measuring bias is race neutrality conditioned on observables. The assumption is that teacher's perceptions and expectations are unbiased if they are based only on legitimate observable predictors of performance, such as past grades, test scores, attitudes about school, and beliefs about personal abilities (for example, as measured by a survey). In this case, the benchmark is only conditionally race neutral; if past performance or another of these predictors is correlated with race, the benchmark will be too. Bias is the difference between the actual perception or expectation and the benchmark for neutrality.

This type of bias can be estimated by regressing a teacher's perceptions or expectations on both race and one or more other explanatory variables that one regards as legitimate predictors of performance. The coefficient of that one race then measures the average racial bias among teachers in the sample. This benchmark is probably more appropriate than unconditional race neutrality when considering, for example, whether teachers rely on biased judgments to nominate students for particular curriculum tracks or ability groups. As I show below, it might also be more appropriate for analyzing whether teachers' biases produce self-fulfilling prophecies of poor performance among black students. However, it is not sufficient to distin-

23. Recall that two lines with different slopes can each represent a correlation of one between the variables on the x and y axes. Similarly, teachers' early perceptions or expectations could have a much larger impact on performance for one race than for the other (as represented by a steeper line), even though the correlation between teachers' perceptions or expectations and end-of-year performance is the same for both groups. The possibility of different slopes is explored by Jusim, Eccles, and Mandel (1996), as discussed below.

guish conditional from unconditional race neutrality; the existing literature often makes a further distinction between past performance and future potential.

Race Neutrality Conditioned on Potential

The third type of benchmark—which may or may not equate with either of the two discussed above—is the level of performance that a student could reach at full potential. In this case, bias is found in the perception or estimation of a student's full potential. Full potential equals demonstrated plus latent potential. It is alleged that teachers underestimate the latent potential of blacks more than that of whites.

It is of major concern to African Americans that teachers underestimate the potential of black students, if not necessarily their performance. Consider the following passage from a 1989 report entitled *Visions of a Better Way: A Black Appraisal of American Public Schooling*: "We hold this truth to be self-evident: all black children are capable of learning and achieving. Others who have hesitated, equivocated, or denied this fact have assumed that black children could not master their school-work or have cautioned that blacks were not 'academically oriented.' As a result, they have perpetuated a myth of intellectual inferiority, perhaps genetically based. These falsehoods prop up an inequitable social hierarchy with blacks disproportionately represented at the bottom, and they absolve schools of their fundamental responsibility to educate all children, no matter how deprived."²⁴

An earlier description likewise alleges bias, judged against the benchmark of future potential: "In the middle class white school, student inattention was taken as an indication of teacher need to arouse student interest, but the same behavior in a lower class black school was rationalized as boredom due to limited student attention span. In general, the teachers in the lower class black school were characterized by low expectations for the children and low respect for their ability to learn."²⁵

24. Committee on Policy for Racial Justice (1989), quoted in Miller (1995, p. 203). The task force that produced the report included a number of noted black scholars at major universities, including Sam Lawrence Lightfoot of Harvard (the principal author), James P. Comer of Yale, John Hope Franklin of Duke, and William Julius Wilson, then at the University of Chicago.

25. Leacock (1969), quoted in Brophy and Good (1974, p. 10).

If as they surely must, perceptions of children's intellectual potential affect the setting of goals in both homes and classrooms, teachers and parents who underestimate children's potential will tend to set goals that are too low.²⁶ Such underestimation is undoubtedly a major problem, irrespective of race. A great waste of human potential and much social injustice results from the fact that teachers are not given the incentives and support they need to set, believe in, and skillfully pursue higher goals for all students, and in particular, for African Americans and other stigmatized minorities. The payoff to searching more aggressively for ways to help children learn would surely be higher than most people imagine.

Reliable estimates of bias related to future potential are not possible, because there is no clear basis on which to measure human potential.²⁷ Surveys find that expressed beliefs in the intellectual inferiority of blacks have moderated over the years.²⁸ In the General Social Survey, for example, the percentage of whites responding that blacks have less "in-born ability to learn" fell from 27 percent in 1977 to 10 percent in the 1996.²⁹ There is no way to know the degree to which this reduction is due to changes in beliefs as opposed to changes in social norms. The same survey found in 1990 that when respondents were not constrained to attribute differences to genetic factors, 53 percent agreed that blacks and Hispanics are less intelligent than whites. Indeed, 30 percent of blacks and 35 percent of Hispanics agreed.³⁰

Many experts think that genetic differences are at least partially to blame for existing black-white differences in academic achievement. A 1984 survey questioned 1,020 experts on intelligence, most of them professors and university-based researchers who study testing, psychology, and education.

26. Goals are not determined by teachers' expectations alone. The curricular materials that are handed down to teachers from the administration, as well as students' actual behavior, also matter.

27. See Ford (1996) for a useful discussion of the issue in light of theories of multiple intelligences. In the terminology of the present chapter, those who think that potential is very distinct from performance and that ability is equally distributed among the races will favor an unconditionally racial neutral proxy in place of race neutrality conditional on observables. Those who believe that racial differences in performance are good approximations of racial differences in potential might favor a proxy of race neutrality conditional on observables, perhaps augmented by a positive shift factor for all students.

28. Miller (1995, chapter 8) presents a useful review of trends in surveys regarding beliefs about black intellectual inferiority. He points out that numbers in the Harris poll tend to produce smaller percentages because they ask more directly about whether any black-white difference in intelligence is genetic.

29. Kluegel (1990, pp. 514-15, 517); see also the introduction to this volume.

30. Miller (1995, p. 183), based on Tom Smith (1990, p. 6).

Almost half (46 percent) expressed the opinion that black-white differences in intelligence are at least partially genetic. Fifteen percent said that only environment is responsible, 24 percent regarded the available evidence as insufficient, and 14 percent did not answer; in other words, only 15 percent clearly disagreed.³¹ With expert opinion slanted so strongly in favor of the genetic hypothesis and widespread media attention paid to books like Richard Herrnstein and Charles Murray's *The Bell Curve*, there is little prospect that "rumors of inferiority" will cease or that racial differences in estimates of students' potential will disappear.³²

Writers concerned with bias in the estimation of potential often claim that it leads to self-fulfilling prophecies. Their point is that children would achieve more if teachers and other adults expected that they could. In most cases, it might be more appropriate to describe bias of this type as producing expectations that are "sustaining" of past trends.³³ Such an expectation is likely to block the absorption of new information into a decision process, and thereby to sustain the trend that existed before the new information was received.

Self-Fulfilling Prophecy

A self-fulfilling prophecy occurs when bias in a teacher's expectation of a student's performance affects that performance. Self-fulfilling prophecies may be associated with any of the three types of bias discussed above, but only those associated with the second type—where the benchmark is conditioned on observables—can be well measured. The basic concept was introduced into social science by Robert Merton in 1948, and Robert Rosenthal and Lenore Jacobson's work on the topic sparked a small industry of studies during the 1970s and early 1980s.³⁴ The effect shows up (and sometimes fails to do so) in a wide range of experimental studies with animals and with human subjects.³⁵ Experimental studies in education typically involve the random assignment of students to groups that have been labeled as high or low performing.

31. Miller (1995, pp. 186-87), based on Snyderman and Rothman (1986, 1987).

32. Herrnstein and Murray (1994); Jeff Howard and Ray Hammond, "Rumors of Inferiority," *New Republic*, September 1989.

33. See, for example, Good (1987).

34. Merton (1948); Rosenthal and Jacobson (1968).

35. See Rosenthal (1994) for a review.

The successful instigation of self-fulfilling prophecies by researchers requires that (1) teachers believe false information about students; (2) teachers act on the information in ways that students can perceive; and (3) students respond in ways that confirm the expectation. The effect can fail to appear—and it often does—if any one of these conditions fails.³⁶ In experiments that confirm the effect, groups labeled as high performing outperform those labeled as low performing. A meta-analysis by Mary Lee Smith identifies forty-four estimates of effect sizes for reading scores, with an average of 0.48 standard deviations distinguishing students with high and low labels.³⁷ The average across seventeen effects for mathematics is much smaller, at 0.18. Why the effects should be smaller for mathematics than for reading is unclear. Perhaps mathematics instruction is less flexible, and therefore less affected by teachers' perceptions.

Brophy, a leader since the early 1970s in research on teacher expectations, asserts that on average, teachers' expectations in real classrooms probably make only a small difference to their students' achievement.³⁸ He adds the caveat, however, that teachers who hold rigid expectations and permit these to guide their interactions with students can produce larger effects. It is plausible, but not established in any literature that I have seen, that expectations of black students might be more rigid than those of whites. Moreover, expectation effects might accumulate from year to year. Surprisingly, there appears to be no good evidence on the degree to which expectation effects accumulate. If small effects accumulate, they could make a larger difference over time. In the short run, even a small difference due to expectations could push a score across the boundary between two grade levels, and thereby become consequential.

In naturalistic studies, the magnitude of self-fulfilling prophecy can be estimated as the coefficient on a teacher's expectation measure, in an equa-

36. The most frequent explanation for failure is that the teachers do not believe information about the students. Sometimes teachers figure out the purpose of the experiment. Other times, teachers have their own sources of credible information or have known the students long enough to form opinions before the experiment begins. In a meta-analysis of eighteen experiments in which IQ or a similar measure of ability was the outcome, Raudenbush (1984) shows very clearly that evidence of the effect was primarily found in studies where teachers had no opportunity to form an independent impression of students before the experiment began.

37. Smith (1980). She does not say what percentage of these effect sizes are calculated from the standard deviation of test score levels as opposed to test score gains.

38. Brophy (1985, p. 304). Specifically, he estimates that teachers' expectations make about a 5 percent difference, but he does not say whether he means the difference in gain over a school year or the difference in total achievement (as in the level on a test). Brophy's statement is based on his own review of the literature, where individual studies seldom cover more than a single school year.

tion where the dependent variable is the student's ultimate performance at the end of a school year. Assuming that the estimated effect of the teacher's expectation is not simply a stand-in for omitted variables, the idiosyncratic contribution of the teacher's expectation is the consequence of bias.³⁹ In some cases, teacher biases exist but do not affect actual scores or grades, either because teachers do not act on their biases or because student performance does not respond to the biased actions that teachers take. Finally, it is also important to note that a teacher's perception of current performance and expectation of future performance can differ, one showing bias while the other does not.⁴⁰

Testing for Racial Differences in Expectancy Effects

Lee Jussim, Jacquelyne Eccles, and Stephanie Madon are the only researchers who have tested for racial differences in the impact of teachers' perceptions on students' test scores.⁴¹ They collected teachers' perceptions of current performance, talent, and effort in mathematics for 1,664 sixth graders in October 1982.⁴² They then tested for what they call racial stereotype bias—that is, whether a student's race predicts teachers' perceptions after controlling for background factors, including previous grades, previous test scores, self-perception of mathematical ability, self-reported level

39. This multivariate equation includes controls for predictors of performance, such as past performance and socioeconomic background. Typically, the estimate of self-fulfilling prophecy may tend to be statistically biased upward, because of omitted variables that are positively associated with both teacher expectations and student performance. Hence any findings of this sort must be taken as suggestive, not definitive.

40. It is not unusual, for example, for a teacher to say, "Betsy is doing well now because she is repeating the grade and has seen some of this material before. I don't expect she will do as well for the rest of the year." This teacher might be accurate in the current evaluation of the student, but still biased in the expectation. Or the example might be reversed, with the expectation more positive than the evaluation of current performance. In either case, the expectation might or might not be biased when judged from the perspective of what past performance and attitudes would predict.

41. Jussim, Eccles, and Madon (1996, pp. 350–51). They speculate that the void in the literature stems from the political risk of studying groups that do, in fact, differ. Instead, researchers have tended to focus on experimental studies that assume away differences. See also Jussim (1989) and Jussim and Eccles (1992), which use the same data to study the accuracy of teachers' expectations without emphasizing racial differences.

42. Of the total student sample, seventy-six are African Americans—ideally, a larger number would be preferable for such a study. The data are taken from the Michigan Study of Adolescent Life Transitions, which was not initially designed to study racial differences; for more on the Michigan study, see Wiegfeld and others (1991).

of effort, and self-reported time spent on homework. This is an example of the second type of bias defined above, using a benchmark of racial neutrality conditioned on observables, including past performance. They find no evidence of racial stereotype bias in teachers' perceptions of current performance, talent, or effort for this sample of sixth-graders.⁴³ The coefficient on student race is small and statistically insignificant.⁴⁴

If racial differences in teachers' current perceptions are explained by students' past performance and attitudes, then these perceptions can only be an important source of a *firm* black-white test score gap if they affect blacks and whites differently. This is precisely what Jussim, Eccles, and Madon find when they analyze the effects of teachers' perceptions of performance, talent, and effort in October on mathematics grades and scores on the mathematics section of the Michigan Educational Assessment Program (MEAP) the following spring semester, in May 1983.⁴⁵ For both grades and scores, the estimated impact of teacher perceptions is almost three times as great for African American students as for whites.⁴⁶ Effects are also

43. Although they do not report raw means by race, they do report that the correlations of race with grades and standard test scores were -0.12 and -0.14 , respectively (in both cases, $p < 0.001$), with black students having the lower scores and grades. These correlations are probably smaller than in the typical national sample of (black and white) sixth-graders.

44. Jussim, Eccles, and Madon (1996) perform separate calculations to determine whether the residual variance in teachers' perceptions left unexplained by the background factors is similar for blacks and whites. They find it to be slightly higher for blacks, but by a margin so small as to be inconsequential: "The correlations of ethnicity with the absolute values of the residuals from the models predicting teacher perceptions were 0.06 ($p < 0.05$), 0.07 ($p < 0.05$), and -0.02 (not significant) for performance, talent, and effort, respectively" (p. 355). While two of the three are statistically significant, these suggest only a very small difference in accuracy, and less accuracy for blacks than for whites.

Regarding other effects, they find a small positive relationship between parental education (a proxy for socioeconomic status) and teacher's perception of a student's talent. There are also some small gender effects: Teacher perceptions of performance and effort are higher for girls, after controlling for the factors listed in the text. Hence it appears that teachers relied somewhat on a gender stereotype, although not necessarily a false one.

45. As background variables, the equation to predict scores and grades includes race, mathematics grades from the end of fifth grade, mathematics scores from the end of fifth or beginning of sixth grade, self-perception of ability, self-reported effort at mathematics, self-reported time spent on homework, and indexes of the intrinsic and extrinsic values of mathematics to the student. Interactions of student race with teacher perceptions of effort and talent were tried, but were found to produce strange results, because of collinearity with the interaction for race and performance. The result for performance might best be interpreted as the interaction of race with all aspects of a teacher's perceptions.

46. The effect size for MEAP scores is only 0.14 for whites, but 0.37 for African Americans ($p < 0.001$). This effect size for whites is quite close to that of 0.17 for mathematics achievement scores reported by Smith (1980). For grades, the effect size for African Americans is 0.56 , compared with 0.20 for whites ($p < 0.01$).

larger for girls and for children from low-income families. Further, the effect is cumulative across disadvantages or stigmas: black children from low-income backgrounds experience the effects of both race and income. Teachers' perceptions of student effort do not affect MEAP scores but do affect grades, even though they are not strongly correlated with self-reports of effort.⁴⁷

What might explain racial differences in the consequences of teachers' perceptions? One possibility is that the result is simply a statistical artifact due to omitted variable bias. This seems unlikely.⁴⁸ More plausibly, teachers are less flexible in their expectations of blacks, females, and students from low-income households. Or, as Rhonda Weinstein speculates, "minority status may play a role in the vulnerability with which students respond to teacher expectations. Differences in cultural values (family compared to school) may serve to immunize some children from the impact of teacher views of their performance or alternately to heighten their susceptibility to the dominant viewpoint."⁴⁹ Perhaps the behaviors of both teachers and students are affected by the combination of the student's race and the teacher's perception of performance. These possibilities are addressed below.

Calculations conducted by Jussim after the paper was published use a specification that includes additional interaction terms, including interactions of race with past grades and scores. The effect size for MEAP scores rises from the original 0.37 to 0.58 for African Americans and drops from 0.14 to 0.13 for whites. Moreover, the coefficients on past grades and scores are estimated to be somewhat smaller for African Americans than for whites. Hence it appears that the performance of these black students was more dependent on teachers' current opinions and less anchored in measures of past performance than that of whites. One might speculate that this is because past grades and scores were less accurate measures of the knowledge or potential of the black students than of the whites, but one cannot be sure from the information available.

47. It is not clear what to make of absence of any relationship between self-reports and teachers' perceptions of effort. If teachers' assessments really are grossly inaccurate, this fact could contribute to the disengagement of children who are believed not to be trying when in fact they are. It could also contribute to a lack of challenge for students who are slacking off—even though they might appear to be working hard—and who would work harder if asked.

48. The fact that teacher perceptions are also strong predictors for females and for whites from low-income households makes it more likely that this is a real effect for blacks. Further, since Jussim, Eccles, and Madon (1996) find no unexplained racial differences in the October performance ratings for blacks and whites after controlling for background factors, and only a trivial difference in unexplained variation, it seems unlikely that the ratings have very different interpretations or different implicit scalings for blacks and whites. Still, these results need to be replicated several times in order to be firmly established.

49. Weinstein (1985, p. 344).

Table 8-1. *Spring Standardized Grades and Test Scores in Mathematics and Fall Performance Ratings, Sixth Grade, 1982-83^a*

Measure and race	Fall performance rating				
	1	2	3	4	5
<i>Predicted spring grades</i>					
Blacks	-1.00	-0.57	-0.14	0.28	0.71
Whites	-0.43	-0.25	-0.07	0.11	0.28
Difference	-0.57	-0.32	-0.07	0.18	0.43
<i>Predicted spring scores</i>					
Blacks	-0.79	-0.46	-0.13	0.20	0.53
Whites	-0.30	-0.17	-0.04	0.09	0.21
Difference	-0.50	-0.29	-0.09	0.11	0.31

Source: Author's calculations based on data from Jussim, Eccles, and Madon (1996, pp. 308-11).
 a. All other student characteristics are held constant. Grades and test scores relate to the mathematics section of the Michigan Educational Assessment Program. Fall ratings are such that 1 denotes the lowest level of current performance and 5 denotes the highest level. The overall mean is zero and the standard deviation is one.

Table 8-1 shows simulated mathematics scores and grades and teachers' performance ratings, holding all other student characteristics constant. For both blacks and whites, there is a positive relationship between teachers' performance ratings in October and the students' grades and scores in May. However, the effect is stronger for blacks. Blacks who receive the highest performance rating (5) in October are predicted to outperform whites who receive that rating. Conversely, blacks who receive the lowest rating in October lag an estimated half standard deviation behind whites with that rating.

If teachers tend to be accurate both in current perceptions and in expectations of future progress, the findings of Jussim, Eccles, and Madon require that teachers expect the pattern shown in table 8-1.⁵⁰ This would

50. Jussim, Eccles, and Madon (1996) assume that current perceptions are good estimates of expectations for the future and they use "perceptions" and "expectations" interchangeably. Jussim also argued in personal communication that this is a reasonable assumption, based on other research regarding the processes by which people form expectations. It might, however, be inappropriate in the present context, for the reasons explained in the text. I do not know of any research that generates data on teachers' perceptions of current performance along with expectations for future performance for both blacks and whites.

represent stereotype bias for expected progress, even if there is no such bias in the evaluation of October performance. The accuracy of the stereotype might reflect self-fulfilling prophecy in the teacher's expectation, or it might not. Evidence that teacher perceptions affect subsequent performance more for blacks than for whites suggests either that black students respond differently than whites to similar treatment from teachers, or that teachers treat black and white students differently, or both.

Do Black and White Children Respond to Teachers Differently?

The finding that black students respond more strongly to teachers' beliefs has not been replicated, but it is consistent with findings from several other studies that ask related questions. In one recent study, Clifton Gasteel asks eighth- and ninth graders whom they most want to please with their classroom.⁵¹ "Teachers" is the answer of 81 percent of black females, 62 percent of black males, 28 percent of white females, and 32 percent of white males. Whites are more concerned with pleasing parents. Doris Entwisle and Karl Alexander find that teachers' ratings of first graders' maturity have larger effects for blacks than for whites on both verbal and arithmetic scores on the California Achievement Test (CAT).⁵² Judith Kleinfield finds that high-school students' perceptions of their own ability are more correlated with perceived teacher ratings of ability for blacks, but more correlated with perceived parent ratings for whites.⁵³ Irvine reaches similar conclusions.⁵⁴

Jussim, Eccles, and Madon suggest, and I agree, that Claude Steele's recent work offers one reason why black and white students might respond differently despite identical classroom conditions.⁵⁵ Steele identifies a phenomenon that he calls stereotype threat, and the resulting stereotype anxiety, that can affect members of any stigmatized group. When the stereotype concerns ability, "threatened" individuals are anxious not to perform in

51. Gasteel (1997). The sample includes 928 whites and 761 African Americans, from twelve classes in nine schools in two public school districts.

52. Entwisle and Alexander (1988).

53. Kleinfield (1972).

54. See the discussion in Irvine (1990, pp. 46-49). In support of her conclusion that "researchers have found that black and other minority pupils are more negatively affected by teacher expectations than white students are," Irvine cites Baker (1973), Krupczak (1972), and Yee (1968). I have not found any studies that contradict these few.

55. Claude Steele, "Race and the Schooling of Black Americans," *Atlantic Monthly*, April 1992, pp. 68-78; see also chapter 11 below.

ways that might corroborate the stereotype. They fear that the stereotype might become the basis of pejorative judgments by others, as well as of their own self-perceptions.

One effect of this anxiety is "a disruptive apprehension" that can interfere with performance. Under stressful test conditions, Steele finds that women and blacks perform worse when they are primed to be conscious of their race or gender. Steele theorizes that when the anxiety is sufficiently intense, it can provoke a response of "disidentification" with the task at hand or with the general category of tasks. Students decide not to consider performance in that particular domain as important to personal goals or self-perceptions.

Steele has tested this idea only for high-achieving students at highly selective colleges. The degree to which stereotype threat and anxiety might apply to students in primary and secondary schools remains to be investigated. Jussim, Eccles, and Madon's findings were for sixth-graders. Are children this young aware enough of stereotypes to be susceptible to stereotype threat or stereotype anxiety? Perhaps.⁵⁶

Susan Gross studied the mathematics performance of students in a racially integrated suburb of Washington, D. C., during the 1985-86 school year.⁵⁷ In fourth grade, 92 percent of blacks and 86 percent of whites who were above grade level on the number of mathematics competencies that they had mastered scored in the eighth and ninth stanines (that is, ninths) of the California Achievement Test for Math (CATM). In the sixth grade, 82 percent of whites who were above grade level in completion of competencies were in the eighth and ninth stanines on the CATM. For blacks, however, the figure was only 68 percent.⁵⁸ Gross points out that for the sixth-graders, this pattern of performance on the CATM was inconsistent

56. As informal but reliable evidence, I offer my personal experience. As a fifth-grader, I moved from a segregated school to one that was integrated. In my new classroom, the top reading group was white, with one or two exceptions; the middle group was mixed, but mostly black; and the slow group was all black. While I did not believe that this pattern was unfair, I wanted the teacher to know, and I wanted to know for myself that I was an exception. The teacher placed me in the middle group. I could not understand why she could not see from my records that I belonged in the top group, despite the fact that I was black. I recall being driven to establish myself as an exception to the racial pattern in the classroom and fearing for a while that my performance in the middle group might not be good enough to do so. I might be trapped in the middle group! After a few weeks the teacher moved me up, and my anxiety abated. However, my constant awareness of racial patterns in group memberships remained.

57. Gross (1993).
58. These fourth and sixth grade results are from a single year for two different cohorts, hence while the differences appear consistent with a trend, they do not clearly establish one. Also, Gross does not report sample sizes broken down into those who were above, at, or below a given grade level. Hence it is not clear how many children these percentages represent.

with in-school performance, and hence she cautions against basing ability group placements on test scores alone.

Gross and her team also conducted focus groups with middle school and high school students. She reports "a deep commitment on the part of high-achieving black students to do well in mathematics so that they could move on to good colleges and professional careers." But the same students felt "deep frustration at the incidents of racism they had experienced in the lower expectations they had perceived from teachers and other students. . . . This was particularly true regarding the honors-level black students who reported that each year they had to prove they were capable of doing honors work."⁵⁹ Moreover, it was common knowledge that children in upper elementary grades felt the same types of pressure.⁶⁰ If the CATM was regarded as a test of ability for sixth-graders, Steele's theory could well explain why black students who were above grade level on competencies got lower CATM scores than their white peers.⁶¹

Gross appears to find evidence of the type of disengagement that Steele hypothesizes, which could help to explain the larger negative impact on black students that Jussim, Eccles, and Madon find when their performance is perceived to be low. Teachers told Gross that black students were overrepresented among students who were "least studious" and "did not come to their classes prepared to work or in the proper frame of mind to attend fully to instruction."⁶²

Both teachers and administrators reported that black parents were less supportive of the school's mission than white parents. However, when Gross convened parents in focus groups, black parents were more supportive of the idea that their children should strive for the higher level classes in mathematics, even if that meant getting lower grades. White parents tended to say that their children should stay in the top sections only if they were likely to do well. Possibly the black parents were sending a mixed message: "Shoot for the top, but if you don't do as well as the white kids, we'll understand."⁶³ If black children sense more ambivalence from their parents than do white children, their teachers' opinions might take on a special significance for them, as the statistical evidence appears to show.

59. Gross (1993, p. 282).

60. Susan Gross, personal communication.

61. Gross reports, in personal communication, that in a regression analysis using students' classroom performance as a predictor, she found that the CATM test scores of the black high-achievers were below those predicted. These calculations were not published, however, and are no longer available.

62. Gross (1993, p. 281).

63. And this, in turn, might have contributed to the ambivalence that seems to be expressed in the work habits of the black schoolchildren, as shown in figure 8-1 below.

In a study inspired by the work of John Ogbu, Roslyn Mickelson distinguishes "abstract" from "concrete" attitudes in twelfth grade students.⁶⁴ She finds that concrete attitudes predict cumulative high school grade point averages, but abstract attitudes do not.⁶⁵ Her measure of abstract attitudes reflects mainstream ideology, standard optimistic rhetoric about education and the American dream. Respondents were asked to indicate their level of agreement with the following statements:

- Education is the key to success in the future.
 - If everyone in America gets a good education, we can end poverty.
 - Achievement and effort in school lead to job success later on.
 - The way for poor people to become middle class is for them to get a good education.
 - School success is not necessarily a clear path to a better life.
 - Getting a good education is a practical road to success for a young black [white] man [woman] like me.
 - Young white [black] women [men] like me have a chance of making it if we do well in school.
 - Education really pays off in the future for young black [white] men [women] like me.
- In contrast, her measure of concrete attitudes includes questions that elicit doubt and ambivalence about education as a route to mainstream success:⁶⁶
- Based on their experiences, my parents say people like us are not always paid or promoted according to our education.
 - All I need to learn for my future is to read, write and make change.

64. Ogbu (1978, 1983, 1987); Herdman and Ogbu (1986); Mickelson (1990).

65. Mickelson studied 1,193 seniors from eight high schools in the Los Angeles area during the 1982-83 school year. She analyzes only the responses of blacks and whites, who compose 41 and 59 percent, respectively, of her working sample. In predicting grade point averages, the standardized coefficients on concrete attitudes in the full specification are 0.111 ($p < 0.05$) for blacks and 0.190 ($p < 0.01$) for whites. These are not large effects. Still, the facts that such a rough index of beliefs is statistically significant at all, and that the distinction between abstract and concrete attitudes is demonstrated so clearly (see table 8-2), are important. The coefficient for abstract attitudes was about a fifth as large as that for concrete attitudes. The t statistic is about 1 for whites and less than 0.5 for blacks. The regressions control for mothers' and fathers' occupation and education, a locus of control index, students' weekly hours worked, percentage of close friends planning to attend a four-year college, and an indicator variable for each of the eight schools in the sample. Mickelson notes that regressions using standardized test scores as the dependent variable produce the same story as for grades.

66. I think that "concrete" is a gross misnomer for these attitudes, and that it has confused the interpretation of Mickelson's work. They are just as abstract as the others: "fairness" is not concrete. They actually measure ambivalence or doubt about the "abstract" attitudes.

Table 8-2. *Abstract and Concrete Beliefs about the Importance of Education for Success, by Race and Socioeconomic Background, Twelfth Grade, Spring 1983^a*

Socioeconomic status and measure	Males		Females	
	Black	White	Black	White
		Effect size		Effect size
<i>White collar</i>				
Abstract beliefs	5.50	0.58	5.27	5.09
Concrete beliefs	4.38	-0.53	4.43	5.00
Sample size	56	...	84	241
<i>Blue collar</i>				
Abstract beliefs	5.28	0.38	5.34	5.21
Concrete beliefs	4.19	-0.36	4.19	4.81
Sample size	138	100	140	93

Source: Mickelson (1990).

a. Sample is taken from eight Los Angeles high schools. Classification as white collar or blue collar is based on a combination of standard blue collar-white collar distinctions and parental education. On each index, higher values denote higher levels of agreement with the given abstract or concrete beliefs; see text for details. Racial differences in abstract scores are significant with $p < 0.05$; class differences for concrete scores are significant with $p < 0.0005$; racial differences for concrete scores are significant with $p < 0.0001$. Full sample standard deviations are 0.76 for abstract scores and 0.98 for concrete scores. Each effect size equals the black-white gap in the respective measure divided by the full-sample standard deviation for either abstract or concrete beliefs, whichever applies.

—Although my parents tell me to get a good education to get a good job, they face barriers to job success.

—When our teachers give us homework, my friends never think of doing it.

—People in my family haven't been treated fairly at work, no matter how much education they have.

—Studying in school rarely pays off later with good jobs.

Students might acquire such concrete attitudes from routine, informal, personal interaction with friends, parents, and other adults, as well as from the broader society.

Mickelson finds that blacks express greater agreement than do whites with the optimistic but abstract beliefs about success and the American dream. However, in their concrete attitudes, which are actually expectations for fairness, blacks are less hopeful than whites. Table 8-2 summarizes the pattern (on both indexes, higher values correspond to higher levels of

agreement). This finding suggests why surveys usually find that blacks subscribe to mainstream values as much as do whites, but behave in ways that show less commitment to mainstream success.

Do Teachers Treat Black and White Students Differently?

I know of only four experimental studies that deal with teachers' treatment of black and white students, all dating from the 1970s.⁶⁷ These studies control differences between students by matching or random assignment. As with most of the experimental literature already discussed, the experiments are contrived one-time encounters. All four experiments find that teachers are less supportive of black than white students.

In Marylee Taylor's experiment, for example, a six-year-old student was said to be watching from behind a screen as college students of education undergoing teacher training taught a prescribed lesson.⁶⁸ In each case, the phantom student was described as black or white, male or female, and of high or low ability. The teachers were told that the students could see and hear them and would respond to their instructions by pushing buttons to activate ten lights on a panel. In fact, all of the "student feedback" was provided by a single adult, who was blind to the description of the "student" given to any particular teacher. Taylor finds that black phantom students received briefer feedback after mistakes (the standardized effect size is 0.613), less positive feedback after correct responses (0.423), and fewer "helpful slips of the tongue"—that is, unauthorized coaching (0.536). Each of the experimental studies suggests that some teachers may help white students more than blacks, and that the differences may be large enough to have nontrivial effects on performance.

67. Coates (1972), Rubovits and Machr (1973), Feldman and Orchowsky (1979), Taylor (1979). See also Bahad (1980, 1985) for related research from Israel, where teachers in the experiment grade assignments on which the names are randomly either European or Moroccan.

68. Taylor (1979) designed her experiment in response to the possibility that confederate students' behaviors had confounded the findings of racial bias in Rubovits and Machr (1973). By placing the phantom student behind a one-way glass, where they could allegedly see the teacher and respond to the teacher's questions and instruction, she removed any effects of targets' actual behaviors. Sessions were videotaped and the teachers' behaviors were coded. The participants were 105 white female undergraduates at the University of Massachusetts, Amherst. They were told that the purpose of the experiment was "to examine certain aspects of teaching behavior in a situation where feedback from pupil to teacher was limited." Discussions after the experiment show that they believed this premise.

Studies of real classrooms confirm this hypothesis. While some find no racial differences, more find differences favoring white students.⁶⁹ The studies that do find differences are probably more likely to be published. Nonetheless, if the benchmark is unconditional race neutrality, there is strong evidence of racial bias in how teachers treat students. It is nearly impossible in naturalistic studies to determine whether teachers would appear racially biased if one controlled racial differences in students' behaviors, work habits, and social skills. But since students and parents cannot read a teacher's mind, they may *think* that differences in treatment reflect favoritism. And when teachers appear biased, trust may be eroded and relationships spoiled.

Evidence on Possible Reasons for Differential Treatment

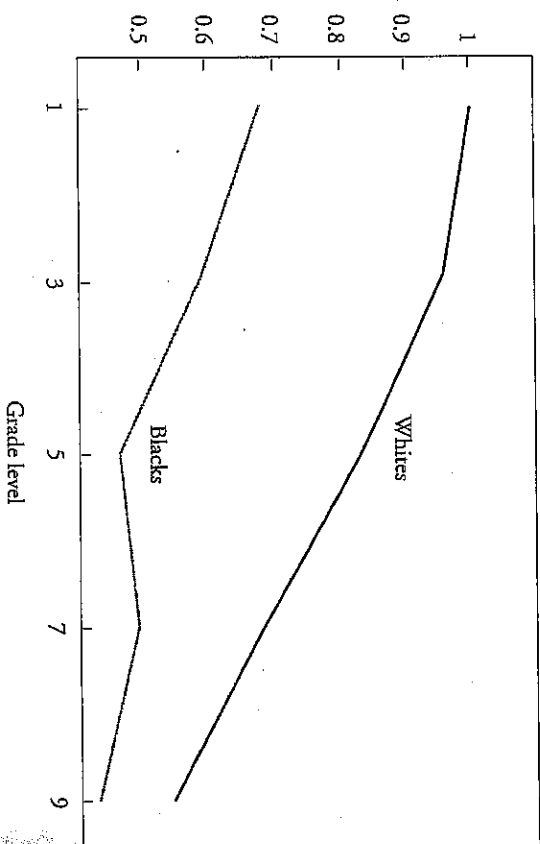
In general, there is no way of knowing whether teachers' perceptions of students' attitudes and behaviors are accurate. Jussim, Eccles, and Madon's finding that teacher perceptions of effort predict grades but not standardized test scores suggests that teachers' perceptions of effort may not be accurate, as does their finding that teachers' perceptions of effort were only moderately correlated with students' self-reports.

In 1990 the Prospects survey, conducted for the Department of Education, asked a national sample of teachers to rate specific students in their classes on the following criteria: "cares about doing well," "gets along with teachers," and "works hard at school."⁷⁰ The three response categories ranged from "very much" to "not at all." Teachers rated black children lower than whites by a statistically significant margin on all three items. To measure the racial differences in these ratings, I subtract for each racial group the percentage with the lowest ranking from the percentage with the highest ranking in each grade, and then sum the results for the three questions. Higher values of the index are therefore better. The index is standardized to equal 1 for white first-graders.

69. There are many sources from which to draw this standard finding. Irvine (1990, table 3.3) replicates studies conducted in naturalistic classroom settings. Of seventeen findings from sixteen studies, whites are favored in nine, there is no difference in four, the opposite race to the teacher is favored in two, the same race as the teacher is favored in one, and blacks are favored in one. One of these studies is from 1969 and the rest are from the 1970s and early 1980s. For discussion of earlier studies, see Brophy and Good (1974).

70. See Puma and others (1993).

Figure 8-1. *Teachers' Perceptions of Students' Levels of Effort^a*
Index^b



Source: Author's calculations based on data collected in the early 1990s for the Prospects study of educational growth and opportunity; see Puma and others (1993).

a. Teachers rated students on the following criteria: cares about doing well, gets along with teachers, and works hard at school. Data for first-graders represent one cohort (1,979 black; 5,333 white); third- and fifth-graders represent a second cohort (1,276 black; 4,045 white); seventh- and ninth-graders represent a third cohort (393 black; 2,424 white). Numbers are weighted to be nationally representative. Chi-square tests on each item for each grade show that through the seventh grade, all black-white differences are statistically significant at the 0.01 level or better. For the ninth grade, the differences are not statistically significant, but the black sample is quite small.

b. Composite index for black and white students in first through ninth grades, constructed such that white first-graders are equal to 1 and higher index values are better; see text for details.

Figure 8-1 shows the means for blacks and whites by grade level. As can be seen, teachers perceive the greatest difference between black and white students in the early years of elementary school. After the fifth grade the gap narrows, but it does not completely close. The apparent similarity in how blacks and whites relate to school by ninth grade is consistent with data on eighth graders from the 1988 National Education Longitudinal Study (NELS). Several authors have remarked on the similarity of black and white attitudes in the NELS.⁷¹ But because that survey did not reach

71. See, for example, Solórzano (1992); Miller (1995); chapter 10 below.

students earlier in their school careers (and also because it did not ask the right questions), it may lead people to underestimate the degree to which racial differences in work habits, behavior, and attitudes below the eighth grade affect teacher-student relations and the black-white test score gap.

Reciprocal Effects between Student Behavior and Teacher Performance

Teachers' judgments about how much they enjoy teaching students inevitably affect their own behaviors. This can apply in regard to entire classrooms and in regard to individual students. Teachers may respond to difficult students by withdrawing support.⁷² This could help account for Jussim, Eccles, and Madoni's finding that teachers' perceptions at the beginning of the school year, although unbiased, are stronger predictors of end-of-year performance for black students than for whites. Specifically, among equally low-performing black and white students, black students might be perceived as more difficult, and therefore receive less teacher support than whites; while among equally high-performing students, black students might be perceived as less difficult, and therefore receive more teacher support than whites.⁷³ I have not found any quantitative research in naturalistic settings that controls for initial student performance and systematically measures racial differences in how much positive reinforcement students provide to teachers. However, on average black students may give teachers less positive reinforcement than do white students with similar levels of beginning-of-year performance.

For example, Willis and Brophy asked twenty-eight first grade teachers to nominate three children to each of four groups, with the following results:⁷⁴

Attachment: If you could keep one student another year for the sheer joy of it, whom would you pick?

Regarding boys in the *attachment groups*, the teachers made more positive comments about their clothing . . . more often assigned them as leaders or

72. On the types of student that teachers like to teach, see Brophy and Good (1974) and the literature that they discuss. Also, see below in the present chapter.

73. Teachers' expectations might be less flexible for black students than for whites. Inflexible perceptions, in turn, might lead to teacher behaviors that reinforce problem behaviors in low-performing students and promote good behaviors in high-performing students. This is a ripe topic for future research.

74. Willis and Brophy (1974, p. 132). Extract is from Brophy and Good (1974, pp. 186-93).

classroom helpers . . . high ability [student] who is well-adjusted to the school situation, conforms to the teacher's rules and "rewards" the teacher by being somewhat dependent upon her and by doing well in his schoolwork.

Indifference: If a parent were to drop in unannounced for a conference, whose child would you be least prepared to talk about?

Boys in the *indifference group* were described as more likely to . . . have a "blank" eye expression . . . to have a disinterested or uncooperative parent . . . to have failed to live up to the teachers' initial expectations . . . Nevertheless, the Metropolitan Readiness Test scores of these boys did not differ significantly from those of their classmates.

Concern: If you could devote all your attention to a child who concerned you a great deal, whom would you pick?

Boys in the *concern group* were especially likely to be described as . . . having a speech impediment . . . being active and vivacious, seeking teacher attention . . . needing readiness work, having generally poor oral and verbal skills . . . and having generally low abilities. . . . [These children are] perceived as making legitimate demands because they generally conform to classroom rules but are in need of help due to low ability.

Rejection: If your class was to be reduced by one child, whom would you be relieved to have removed?

Boys in the *rejection group* were described as being more likely to be non-white than white, coming from intact families in which both parents were living, as being immature and not well-adjusted, as being independent, as being loud or disruptive in the classroom, as being rarely inactive or not vivacious . . . as needing extra help because of generally low ability, as needing readiness work. . . . These children did not differ significantly from their classmates on the Metropolitan Readiness Test scores despite the teachers' comments about low ability.

The rejection group is the only one in which nonwhite boys are over-represented in teachers' remarks. Clearly, much more was involved in shaping the teacher-to-student relationship than simply the child's initial ability or academic performance, at least in first grade. Children's work habits and behaviors (and sometimes even their parents' behaviors) affected teacher preferences. Figure 8-1 has shown that in the 1990s teachers perceive that blacks rate lower than whites on attitudes, effort, and behavior. Based on these patterns, my guess is that *on average* teachers probably prefer to teach

whites, and *on average* they probably give whites more plentiful and unambiguous support.

Mismatches of race between teachers and students do not appear to be the central problem. Even black teachers need help in learning to cope with some of the special demands that black children from disadvantaged backgrounds may present.⁷⁵ Paula, a young black teacher enrolled in a program to help teachers understand, manage, and teach difficult students, admitted:

The first thing I knew was that they were just BADD. I know part of the problem was myself because I was saying things that I probably shouldn't have said because they got me so upset and I wasn't able to handle it. . . . I felt that being black I would automatically know more, and so forth, and in ways I think I do, but [the training program] has helped me to understand things from many perspectives. . . . Black teachers who have been in different programs . . . haven't got this cultural awareness and I know that because they're so negative. . . . A lot of them aren't culturally sensitive to their own culture.⁷⁶

It remains an open question how much difference teachers' preferences about whom they teach make to the educational outcomes of students. In difficult schools there may be many "burned out" teachers, who are simply going through the motions and waiting for retirement. It is also unclear to what degree this pattern of bad student behavior and teacher burnout is racially distinct. In many classrooms, teachers and students are embroiled in conflict and confusion that they lack the skills and external support to resolve. Research to analyze the effectiveness and replicability of programs such as that in which Paula was enrolled should be a priority in order to improve the schooling of black children in settings where behavior is a problem.

Indeed, signals about performance can have racial overtones, and these can interfere with teacher-student relations and with learning. A summary of focus group discussions with black males between the ages of nine and nineteen in Fort Wayne, Ind., in 1996 reports: "Students expressed disappointment in their grades when moving from one class to another, but could not explain the differences from year to year with the exception of

⁷⁵ Simply matching the race of the student and the teacher is too simple a prescription: social class and professional competence also appear to be important. In chapter 9 below, I discuss how a teacher's race and social class might affect student performance.

⁷⁶ Caballo and Burslein (1995, pp. 289-90).

saying that the teacher was prejudiced. Racial prejudice of the teachers was a perception that was common in all the groups. . . . The teacher who encouraged and expected more from the students was always mentioned, but only as the exception. "77 Teachers in integrated schools can be "biased" in ways as simple as reinforcing a propensity of white children to speak more often in class.⁷⁸ As a result, black students may assume that the teachers think whites are smarter or like the white students better. How teachers communicate about academic ability—especially in integrated schools where the performance of whites is superior—can affect the degree to which black students disengage from the pursuit of excellence, or alternatively, stay engaged and aim for mastery.⁷⁹

Race and the Role of Peer Culture

What teachers communicate to students about ability is important, because positioning in the hierarchy of *perceived ability* has social significance for both individuals and groups—and this, in turn, has feedback effects on school performance.⁸⁰ Readily observable racial patterns in messages from teachers or black-white differences in actual performance can produce stereotype anxiety and can blur the distinction between racial and achievement-related aspects of students' identities.

Shared stereotype anxiety probably encourages black students to form peer groups that disengage from academic competition. Members may secretly want to *be* "smart," like the students (often disproportionately white) whom teachers favor. Nevertheless, they may also resent any effort by black peers to defect from group norms by *acting* smart. In one school district with which I am currently familiar, it is common to hear that some black students accuse their peers of "acting white" when their personal styles seem

77. Jones (1997, p. 9).

78. See, for example, Katz (1973), cited in both Brophy and Good (1974) and Irvine (1990).

79. In particular, it is important that teachers make clear that academic ability is not immutable. That sustained effort mobilizes and develops ability. Dweck (1991) and Dweck and Leggett (1988) find statistically significant evidence among white children that those who believe that ability is fixed tend to adopt performance goals, whereas those who believe otherwise tend to adopt mastery goals. Among those who believe ability is fixed and that their own endowment is low, the performance goal serves to hide their ignorance. But both high and low achievers among those who believe that ability can be developed tend toward mastery goals. The same is probably true for blacks, but I do not know of any research that addresses this issue.

80. See the literature review in chapter 10 below.

to resemble those of the smart white kids. However, the same black students who make the accusations resent any insinuation that they themselves are stupid. There is duality. It seems that being smart is valued, but acting smart—or aspiring to move up in the achievement hierarchy, with the associated peer and teacher relationship patterns—is frowned on, at least in others.

How much these factors help to account for the black-white test score gap is uncertain. Philip Cqok and Jens Ludwig conclude in chapter 10 below that adolescent peer pressure and the acting white hypothesis bear virtually no responsibility for black-white differences in academic performance. Their study advances the debate, but it leaves the central question unresolved, as I explain in a comment following that chapter. Nevertheless, rather than simply attacking black peer culture, reducing the amount of unresponsive and ineffective teaching is almost surely a more important response to the black-white test score gap.

Responsive Teaching

As noted at the outset of this chapter, the average black child arrives at kindergarten with fewer academic skills than the average white child. Schools may then push students along in ways that sustain or add to racial disparities, validating the expectation that black-white differences in achievement are normal, perhaps even inevitable. But if instruction is appropriately stimulating and responsive to children's progress, teachers' expectations may be neither self-fulfilling nor sustaining. The more inviting and responsive instruction is to children's own efforts to improve, the less teachers' initial perceptions and expectations will predict later success.

Research that measures how instructional methods affect the accuracy of teacher expectations is rare. One relevant set of studies deals with "wait time"—that is, how long a teacher waits for students to raise their hands, for students to begin talking, and for students to continue talking after a pause. Minority students in integrated classrooms participate more when wait time is longer. This improves their performance relative to whites and changes teacher expectations.⁸¹

81. In a summary of the literature on wait time, Rowe (1986, p. 45) reports that teachers' "expectations change gradually, often signaled by remarks such as 'He never contributed like that before. Maybe he has a special 'thing' for this topic. . . . This effect was particularly pronounced where minority students were concerned [emphasis added]. They did more task relevant talking and took a more active

Corrective feedback is probably more significant than wait time, however. In a study that does not mention race, Thomas Guskey looks at forty-four intermediate and high school teachers who taught various subjects in two metropolitan school systems.⁸² Each teacher taught two matched classes, after receiving training. One class was instructed using the teacher's standard methods, and the other was taught with a "feedback and corrective" process learned in the training. Both classes received the same final examination and grading standards. Guskey compared teacher ratings of "probable achievement" from early in the semester with final grades and examination scores, and also with end-of-term ratings of "achievement potential."⁸³ For ten teachers, the training made no difference to their students' performance. However, for thirty-four others—the "positive change" group—the experimental classes did better than the controls on both grades and scores. Among this group, teachers' early expectations were less predictive of students' later achievement as a result of the improved techniques for feedback and correction. Specifically, as shown in table 8-3, correlations between teachers' initial ratings of probable achievement and the students' final grades and examination scores were markedly lower for experimental classes than for classes using customary methods; "no-change" teachers had high correlations in both classes. It seems likely that better feedback and corrective methods could also affect the rank order of performance by race, although Guskey does not investigate this issue.⁸⁴

It is worth noting that responsive teaching can take negative as well as positive forms. For example, some teachers may give incentives and assistance to students who want to improve their positions in the class and penalize students who do not. In an often cited study, Karen Bratesani, Rhonda Weinstein, and Fernine Marshall compare fourth, fifth, and sixth grade classes in which student surveys had indicated various levels of differential treatment.⁸⁵ In the classrooms with higher levels of differential

part in discussions than they had before." Rowe also makes other references to the studies that develop these findings. Wait time is shorter for low-performing students (see chapter 8).

82. Guskey (1982).

83. For the ratings on probable achievement in the course and achievement potential, teachers assigned each student to one of five groups of equal size.

84. Studies find mixed results regarding techniques to improve corrective feedback. See, for example, Slavin's (1987) review of the literature on mastery learning.

85. Bratesani, Weinstein, and Marshall (1984). Their point in this paper is that teacher expectations become self-fulfilling prophecies only when communicated through differential treatment. They find that teacher expectations are stronger predictors in classrooms with higher levels of differential

Table 8-3. *Effects of Improved Techniques in Feedback and Correction^a*

	<i>Correlation between initial rating and</i>			
	<i>Final rating</i>	<i>Course grade</i>	<i>Final examination</i>	
<i>Teacher sample</i>	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>
Positive change	0.53	0.83	0.51	0.80
No change	0.92	0.90	0.77	0.79
			0.79	0.69
				0.75

Source: Guskey (1982).

a. Sample comprises forty-four intermediate and high school teachers of various subjects in two metropolitan school systems; see text for details. The positive change group includes the thirty-four whose classes taught using the improved feedback techniques earned higher final examination scores and course grades than their control classes. For the ten teachers in the no change group, either course grades or scores on the final examination were higher in the control classes.

treatment, more students with below average scores at the beginning of the year made unusually large gains, but fewer students with above average scores made gains.⁸⁶

Both of these studies show that greater responsiveness to individual children can weaken the link between past and future performance, and perhaps also alter trajectories. Both are silent about race and ethnicity. Unfortunately, statistical studies that deal directly with race do not investigate whether particular teaching practices can change the rank order of performance among students.

The Great Expectations Initiative

Great Expectations is a public-private partnership created in 1989 to bring Marva Collins's ideas about teaching into Oklahoma schools.⁸⁷

treatment. However, since they collected teachers' expectations in April of the school year, I would regard these as reports from the end of the school year rather than as self-fulfilling predictions. For interesting related work, see Weinstein and others (1987), who show that even first-graders can accurately report teachers' differential treatment of their peers, but that it is not until third grade that students begin to give accurate reports regarding their own differential treatment.

86. See Bratesani, Weinstein, and Marshall (1984, table 4). They find strong evidence that standardized test scores are predicted less well by past performance in classrooms where there is more differential treatment. The sample sizes are small, so that the difference in percentages of low achievers who make large gains does not reach statistical significance. Nevertheless, the magnitudes of the gains are large.

87. This summary draws from Ferguson (1993), which tells the story of the birth and early development of the initiative.

(Collins, an African American, teaches in the inner city in Chicago and is probably the most widely known elementary teacher in the nation.) The initiative includes a range of techniques that Collins has developed over the years. It aims to nurture in all students, not only the most talented, the expectation that they are destined to be important people if they do their best in school to prepare well for the future. Those who misbehave should be reminded regularly that the teacher cares and refuses to give up on them. Teaching methods combine high challenge for students with feedback from teachers *and peers* in forms that make learning fun and emphasize its importance for a happy and effective life. Progress is celebrated, so that every student can earn the opportunity for positive recognition from teachers, peers, and parents. In addition to more standard materials for core subjects, the curriculum includes uplifting, forward-looking poetry, which students memorize and discuss and recite at school and at home. The story of Great Expectations shows real people struggling, with some success, to change teaching practices—and in the process, teachers' expectations—for disadvantaged, mostly minority, children. Whites, blacks, Hispanics, and native Americans are all represented among the poorly performing children that the initiative aims to help. Racial gaps are not an emphasis.

The incentive for the Great Expectations initiative was a threat of takeover of certain Oklahoma schools by the state if test scores for third-graders persisted below the twenty-fifth percentile for three consecutive years on the Iowa Test of Basic Skills. Educators in the schools that joined Great Expectations knew of Marya Collins's reputation for working wonders with children in inner-city Chicago. Her own school had never been independently evaluated, but she appeared to be effective with the types of children that the Oklahoma schools were failing. Although administrators were not certain that these methods could be transferred from Chicago to Oklahoma, they judged it worth a try. For the first training, two teachers from each of twenty-five pilot schools were sent to Collins's Westside Preparatory School in Chicago. There they had a "seeing is believing" experience concerning what children from disadvantaged backgrounds could achieve.

As the initiative spread through the pilot schools in Oklahoma, however, there was substantial resistance from teachers who had not gone to Chicago. The head mentor teacher told me that virtually all of the resistance she encountered represented one or more of the following three perspectives:⁸⁸

88. She maintained that most of the time the resistance could be reduced by a combination of two responses. First, she would assure the teacher that he or she could slip into the Great Expectations

—Time: "I just don't have the time to try this. It's too much. I just can't do it."

—Satisfaction with current practices: "I just don't see the need for doing things differently from what I already do."

—Hopeless students: "You don't know my kids. You couldn't do that with my kids. All that positive stuff is treating kids like babies: discipline has to be tough — you can't mix it with being nice."

Some teachers were insecure: their low expectations for students were partly the consequence of low expectations for themselves as teachers.⁸⁹ At the other extreme, some teachers thrived using Collins's ideas and felt professionally rejuvenated. Each summer during the 1990s several hundred teachers have been trained or have received refresher courses through free summer institutes at Northeastern State University. Through the work of the Great Expectations Foundation, funding for the institutes and mentor teachers comes from private donors, philanthropic foundations, the State Board of Regents, and the state Department of Education. School-site training by mentor teachers during the school year reinforces the training received at the summer institutes. Staffs for the institutes comprise professors from several NSU departments and elementary school teachers who have distinguished themselves in the classroom using the Great Expectations teaching methods.

In a baseline survey administered at the Summer Institute in July 1993, I asked teachers with some previous Great Expectations training to describe the improvement in their students' performance since they began using these methods. Table 8-4 presents the results.

The two portraits that follow offer "existence proofs" of the proposition that teaching practices and expectations can change dramatically, even for experienced teachers. At the same time, both of the teachers described express reservations.

program, gradually, implementing some elements first and others later. Second, she would model the specific practices to which the teacher was resistant. She would do so at that teacher's school, preferably in the teacher's classroom, and always with the greatest respect and tact. When a teacher witnessed a mentor successfully demonstrating a method with the teacher's own students, he or she usually became (or claimed to become) more open to giving it a try.

89. One worst-case example involved a first grade teacher who had failed with a different new method during the previous year. Although several other teachers at her school were using the Great Expectations method and doing well, she was sure that it could not work for her, and she received no pressure to change from the passive principal.

Table 8-4. *Teachers' Assessments of Student Progress due to the Great Expectations Program, Oklahoma^a*

Assessment	Aspect of classroom performance			
	Academic performance	Attitudes	Behaviors	Teacher's job satisfaction
More than I thought was possible	22.37	31.59	25.00	35.53
A lot	55.26	46.05	44.74	48.68
Some	22.37	19.74	28.95	13.16
None	0.00	2.63	1.32	2.16

Source: Author's tabulations from a survey conducted at the Great Expectations Summer Institute, July 1993.

a. Sample comprises seventy-six teachers with some prior training in Great Expectations methods (representing close to a 100 percent response rate of potential interviewees). Participants were asked to complete the following statement: "Because of Great Expectations, the improvement in [aspect of classroom performance] of my students has been. . . ."

Greg Robarts, Fourth Grade Teacher, Beehive Elementary School

Greg Robarts's classroom is a roughly even mix of black, white, Chicano, and Native American children, almost all of whom come from very poor families.⁹⁰ Before the Great Expectations initiative, Robarts had taught for seventeen years and believed himself to be a good teacher. But seeing what children at Westside Preparatory School in Chicago could do gave him pause: "I didn't really know how to teach reading. After one workshop in phonics I feel that I know more today than I learned in seventeen years teaching." He describes seeing Westside Preparatory School as "an awakening." "I saw something I'd never seen before; I actually saw education taking place. I saw children interested in learning. After seeing her approach, and seeing that it worked, I thought, 'What I'm doing now isn't working. At best it's just kind of passing.' . . . I had to rededicate myself."

Collins's basic philosophy resonated with Robarts's beliefs about teaching, but he had to unlearn old habits, such as sitting at his desk saying, "Open the book to page 34, here are the instructions. . . ." Even his own

90. In this and the following subsection, names of teachers and schools have been changed, but facts and quotations are real.

principal described Robarts as a "desk sitter" before he changed to the Great Expectations way of running a classroom: "Teach on your feet, not in your seat." Before Great Expectations,

I was secure with all the books and things. A lot of teachers are where I was. They're embarrassed to say "I don't know." It's that fear of judgment . . . teachers are hesitant to ask . . . Teaching independently . . . instead of from the book, those are the kinds of things that I wasn't courageous enough to try.

[How long was it before you felt comfortable with this new style?]

Oh, I think after about the first day. And I made some horrible mistakes.

But my kids just hugged me and said, "Oh, Mr. Robarts, you're so different from when you left!" And they would just say, "Oh, you're doing well." And when I would start kind of maybe, "Well, maybe I need to sit down now; I've kind of done the Marva Collins stuff now, so maybe I need to sit down." The kids would say, "Mr. Robarts, we sense that you're being average again." And so, I said, "Okay." So I always asked them to encourage me when they sensed that I was being average or substandard.

Many people who are not familiar with the Great Expectations approach say that it overemphasizes memorization and underemphasizes higher order thinking. Robarts disagrees. He says these functions are complements, not substitutes: memory is the foundation for higher order thinking. He finds that without practice at memorization many children cannot remember a dictated sentence long enough to write it down. After a few weeks of memorizing poetry and other things, he says, the change is remarkable. He thinks that people who dismiss memory work as outmoded are simply uninformed about how children learn, not only because memory supports higher order thinking, but because children can memorize things that are worth knowing. In addition, by reciting what they have memorized, children build self-confidence and motivation. He says: "If you had told me two years ago that I would have a class of fourth-graders that would know all the states and capitals and would know geographically where things are located, that could spell words, that could read, that could do . . . I would have said, 'Well, maybe if you're in Quail Creek—which is a very affluent area of Oklahoma City—perhaps. But in this area, no, it wouldn't happen. . . . You know, maybe rote memory is not positive in some aspects, but I think that when a child has experienced total failure all through school it can be a major first step."

Much of the memory work in Great Expectations classrooms involves

poetry that contains rules for healthy and productive living—messages worth remembering. Robarts reports exciting results from his efforts: "absenteeism is almost nil, refusal to do homework is almost nil, test scores are substantially up." He also describes a "miracle turn-around student" during his first semester using Collins's methods. The student's disciplinary folder was a "blizzard of suspensions." An African-American boy diagnosed as learning disabled, he was "a throw-away child," Robarts says. "He was not supposed to be able to do anything. He came very hostile . . . a tough cracker to break. I didn't understand when Collins said, 'You can look in the eyes of the children when they come, and there's a dullness.' I know what she means now. Children like Jerry have been so programmed to believe that the school is nothing. That *they* are nothing, that the only guarantee they have in school is failure. And it's so exciting to see their eyes brighten, and to see the *child* say that they *can* do." When Jerry transferred to a new school the following year, his classmates teased him for speaking standard English. Jerry persisted, with support from his new teacher. On returning to visit Mr. Robarts, Jerry reported that his new classmates began to change to be more like him, because of the positive responses that he was getting from the teacher.

Robarts realizes that he is not typical. Other teachers need more ongoing support. When I first interviewed him, in December 1991—after the first summer institute but nine months before mentor teachers began working in teachers' classrooms—Robarts was sober about the value of a four-hour demonstration that he and his class were to give the next week for teachers who had not gone to any summer institute: "We'll get them excited. They'll go back to their classrooms, they'll meet with the same failures that they've had. They'll struggle. They'll crash. They'll burn. They'll say, 'To hell with it. It's just another thing that they're doing.' And that will be the end of it. If there is no follow-through, no support person, no person to be supportive and say, 'Well now, this is a possibility,' it will all come to naught." In fact, Robarts was among the teachers who had pushed for the establishment of summer institutes and the use of mentor teachers for ongoing technical assistance. Currently, both programs remain in place, complemented by an academy for principals.

Gloria Chavers, Third Grade Teacher, Lafayette Elementary School

Gloria Chavers recalls not thinking much about Great Expectations at the time that her school applied to participate in the program. However, when the first two teachers came back from Chicago, "They were excited,

"There was no doubt about that." The principal asked other teachers to observe in the classrooms of the trained teachers. "So we went to Mrs. Sherrin's [third grade] room. But this excitement that she had, I couldn't pick up on it. Because, and I talked at length with the principal about it at the time, I hadn't experienced what they had experienced. And for them to sit and tell me about what a five-year-old, and they would call these little children's names, you know, they could recite all this. I'd never been around any children who could do this, so it was hard for me to envision."

Chavers also recalls that she saw changes that she did not like in Sherrin's students. She had taught some of these children herself, as a second grade teacher. Now, they were calling themselves "college bound." At that time, Chavers's view was that "in this school system, college is not for everyone. We have a lot of lower socioeconomic people. College is the exception, not the rule."

Finally, the opportunity came to attend the first summer institute. "As it turned out, it was really well worth it," says Chavers. The following semester, she reorganized the way she ran her classroom. "We've gone back to a highly structured way of reading and teaching phonics [using chants]. We'd gotten away from that." She also now teaches the whole class in one group and from one book. She reports that when she first changed, "Some children struggled, but it's surprising when they have their peers reading and reading well, it seems to give them more incentive to read better."

At the summer institute, Chavers learned how to teach addition and subtraction using chants, but the class had not gone on to multiplication and division. So, back at her school, she made up her own chants for multiplication and division. She recalls: "And then I told [the students] one day I said, 'Well, we'll sing this out.' Well, they didn't know what that was. I told them, 'It's like you're standing on the corner trying to sell something.' And even the children who have more difficulty with math, they have been able to pick up on those multiplication tables. They can not only say them, they can pass their tests! A lot of times after we do them, they'll go out of the room and you'll hear them going down the hall buzzing, singing them. You know, they like to do it. It's really, it's not anything new, it's just the way it's presented."

Chavers talks on about bringing her love of music into the classroom now in ways she never felt authorized to do before. She talks about impressing her friends with her students' written work. She speaks with pride about parents who glow when they see report cards that are better than ever before, who brag that their children are doing work that the parents them-

selves did not see until junior high school. Parental interest and participation has clearly increased. According to the district superintendent of a rural white community where one of the schools is located, "Some parents here were kind of skeptical about going up and bringing 'this black thing from Chicago,'" into this white, mostly rural section of Oklahoma. The same parents became supporters, however, when they saw the difference that it made for their children.

Chavers says that her children are convinced that they can do anything. When she plays choral music on a tape recorder they beg to learn the songs: "This week they said, 'Oh, won't you play the music?' And, 'Oh, can't we learn the song?' . . . And they assured me, 'Oh, we can learn it.' So in two afternoons, they pretty well learned it. I was once a music teacher. With this new program I've been able to incorporate it again." When asked who gave her permission to do so, she says, "I just did it. I don't have to feel like this isn't part of my work anymore." Other teachers at other schools expressed similar feelings of a new freedom to bring their personal interests and talents into the classroom.

At the time of our interview, Chavers had been teaching for seventeen years. But, she says, "With the introduction of this program, it's just been different. The whole atmosphere has been different around here. The discipline problems for me have all but just totally disappeared, with this program. And it's not the fact that you're after the kids all the time. It's, 'This is what I expect of you.' You know, 'You are here for a job. This is your job, and my job here is to teach you. Your job is to be the best student you can be. And that is what I expect of you.'"

Robarts and Chavers are examples of what is possible, though perhaps not for all teachers. According to the head mentor teacher, the most important distinction between schools that do very well with Great Expectations and those that do not appears to be having an effective principal who understands the initiative. One characteristic of such principals is that they find ways of removing ineffectual or uncooperative teachers from their schools. The outcomes of the Great Expectations initiative have not yet been rigorously evaluated. However, several teachers bragged during interviews that average test scores in their own classes had risen by 30 or more percentiles in the space of one year.⁹¹ The key for teachers is the apparently

91. Chapter 9 discusses the Success for All program, which serves a large number of mostly black and Hispanic children across several states. It shows positive effects for all students, but larger effects for the bottom 25 percent of the class.

effective program of professional development that has helped them to expect more and achieve more for both themselves and their students.

Conclusion

Any conception of bias requires a corresponding conception of neutrality. A major reason that no consensus has emerged from scholarship concerning the importance of racial bias in the classroom is that there is no single benchmark for racial neutrality. Instead, there are at least three: unconditional race neutrality, race neutrality conditioned on observables (including past performance), and race neutrality conditioned on unobserved potential. Moreover, racial biases can exist in teachers' perceptions, expectations, or behaviors, or in any combination of the three.

Consider teacher perceptions of current student performance. If the benchmark for bias is unconditional race neutrality, most teachers are biased, but evidence shows that this is mainly because their perceptions of current performance are correct. When their perceptions early in a school year are inaccurate, the inaccuracies may become true through a process of self-fulfilling prophecy, but there is little evidence that initial inaccuracies or prophecies systematically favor either blacks or whites. In fact, where the benchmark is racial neutrality after taking past performance and other observable predictors into account, evidence favors the conclusion that teacher perceptions of current performance are generally unbiased. Whether the same applies to expectations and behaviors is less clear. I have found no clear evidence on whether teachers' expectations or behaviors are racially biased for students whom they perceive to be equal on past or present measures of performance or proficiency. However, taking unconditional racial neutrality as the benchmark, it is clear that teachers' perceptions and expectations are biased in favor of whites and that teacher behaviors appear less supportive of blacks. Clearly, the benchmark chosen for neutrality affects the conclusions.

Robert Schuller says, "Any fool can count the seeds in an apple, but only God can count the apples in a seed."⁹² Similarly, tests can measure what children know, but only God can measure their latent future potential. Neutrality conditioned on latent future potential relates to a third type of bias and a third way in which teachers' beliefs can matter. Since potential

92. Robert Schuller is a popular television minister and proponent of positive thinking.

is unobserved, racial bias of this type is virtually impossible to gauge with any reliability. Still, it does seem especially likely that teachers underestimate the potential of students whose current performance is poor, including disproportionate numbers of blacks. Also, blacks are underrepresented among students with the very highest scores, and potential for greater black representation at the top of the distribution is unproven. Thus at both ends of the test score distribution, stereotypes of black intellectual inferiority are reinforced by past and present disparities in performance, and this probably causes teachers to underestimate the potential of black children more than that of whites. If they expect black children to have less potential, teachers are likely to search with less conviction than they should for ways to help these children to improve, and hence miss opportunities to reduce the black-white test score gap.

Simply cajoling teachers to raise their expectations for black children—using phrases such as “All children can learn”—is probably a waste of time. However, good professional development programs can make a difference. Recall that some teachers in Oklahoma responded to the Great Expectations program with the assertion, “My kids couldn’t do that.” If they had gone on teaching as they had always done, that judgment would have been correct. But when they changed their teaching methods, they learned that they were wrong. Similarly, Guskey shows that teachers can learn responsive teaching methods that weaken the link between past and future performance.⁹³ Teachers who have been helped to improve their classroom practices can have “seeing is believing” experiences that challenge their prior biases. More research is needed on how professional development programs affect both test score levels and the black-white test score gap.

Even in the absence of the biases discussed above, teachers’ beliefs probably affect black students more than whites. The evidence is quite thin, but the few studies that bear on this hypothesis appear to support it. Jussim, Eccles, and Madon find that teachers’ perceptions of sixth-graders’ mathematics performance in October do not contain a racial bias once they control past performance and attitudes.⁹⁴ Nevertheless, the effect of teachers’ October perceptions on students’ mathematics scores in May is almost three times larger for blacks than for whites. Further, the effect is also larger for females than for males, and larger for both black and white students from low-income households. Findings from other studies are consistent

with these results. Castrel finds that black eighth- and ninth-graders are more eager to please their teachers, but their white peers are more concerned about pleasing their parents.⁹⁵ These differences may be due to parenting. For example, white parents might exert more consistent pressure for good grades; black parents might be less assertive about grades and more deferential themselves to teachers. Future research should actively pursue these questions, including the implications for policy, teaching, and parenting.

My bottom line conclusion is that teachers’ perceptions, expectations, and behaviors probably do help to sustain, and perhaps even to expand, the black-white test score gap. The magnitude of the effect is uncertain, but it may be quite substantial if effects accumulate from kindergarten through high school. The full story is quite complicated and parts of it currently hang by thin threads of evidence. Much remains on this research agenda.

Fortunately, successful interventions do establish that children of all racial and ethnic groups have more potential than most people have assumed. As the evidence accumulates, it should be possible to focus with greater determination on cultivating and harvesting all that youthful minds embody.⁹⁶ It would then be no surprise if the black-white test score gap began to shrink again, as it did in the 1980s—and ultimately disappeared.

References

- Babod, Elisha Y. 1980. “Expectancy Bias in Scoring as a Function of Ability and Ethnic Labels.” *Psychological Reports* 46: 625–26.
- . 1985. “Some Correlates of Teachers’ Expectancy Bias.” *American Educational Research Journal* 22(Summer): 175–83.
- Baker, S. H. 1973. “Teacher Effectiveness and Social Class as Factors in Teacher Expectancy Effects on Pupils’ Scholastic Achievement.” Ph.D. dissertation, Clark University.
- Baron, Reuben, David Y. H. Tom, and Harris M. Cooper. 1985. “Social Class, Race and Teacher Expectations.” In Jerome B. Dusek, ed., *Teacher Expectancies*. Hillsdale, NJ: Erlbaum.

93. Guskey (1997).

94. In addition to strong leadership and professional development for teachers, better-conceived performance incentives, no matter what their expectations of their students, should be a part of this process. The search for ways to design and implement such incentives and standards of accountability is currently quite active; see, for example, Hanushek (1994); Hanushek and Jorgenson (1996); Ladd (1996).

93. Guskey (1982).

94. Jussim, Eccles, and Madon (1996).

- Bartasani, Karen A., Rhonda S. Weinstein, and Hermine Marshall. 1984. "Student Perceptions of Differential Teacher Treatment as Moderators of Teacher Expectation Effects." *Journal of Educational Psychology* 76: 236-47.
- Brophy, Jerome. 1985. "Teacher-Student Interaction." In Dusek, ed., *Teacher Expectancies*. Brophy, Jerome E., and Thomas L. Good. 1974. *Teacher-Student Relationships: Causes and Consequences*. Holt, Rinehart, and Winston.
- Cabell, Beverly, and Nancy Davis Burstein. 1995. "Examining Teachers' Beliefs about Teaching in Culturally Diverse Classrooms." *Journal of Teacher Education* 46(September-October): 285-94.
- Careel, Clifton. 1997. "Attitudes of African American and Caucasian Eighth Grade Students about Praises, Rewards, and Punishments." *Elementary School Guidance and Counseling* 31(April): 262-72.
- Coates, Brian. 1972. "White Adult Behavior toward Black and White Children." *Child Development* 43: 143-54.
- Committee on Policy for Racial Justice. 1989. *Visions of a Better Way*. Washington: Joint Center for Political Studies.
- DeMeis, Debra K., and Ralph R. Turner. 1978. "Effects of Students' Race, Physical Attractiveness, and Dialect on Teachers' Evaluations." *Contemporary Educational Psychology* 3: 77-86.
- Dweck, Carol. 1991. "Self-Theories and Goals: Their Role in Motivation, Personality and Development." In Richard A. Dienstbier, ed., *Nebraska Symposium on Motivation*, 1990. Lincoln University Press.
- Dweck, Carol, and Ellen L. Leggett. 1988. "A Social Cognitive Approach to Motivation and Personality." *Psychological Review* 95: 256-73.
- Eccles, Jacquelynne, and Allan Wigfield. 1985. "Teacher Expectations and Student Motivation." In Dusek, ed., *Teacher Expectancies*.
- Egan, Owen, and Peter Archer. 1985. "The Accuracy of Teachers' Ratings of Ability: a Regression Model." *American Educational Research Journal* 22: 25-34.
- Entwisle, Doris R., and Karl L. Alexander. 1988. "Factors Affecting Achievement Test Scores and Marks of Black and White First Graders." *Elementary School Journal* 88(5): 449-71.
- Bertrson, Carolyn M., Jerome Brophy, and Thomas L. Good. 1972. "Communication of Teacher Expectations: First Grade," report 91. University of Texas at Austin, Research and Development Center for Teacher Education.
- Feldman, Robert S., and Stanley Orchowsky. 1979. "Race and Performance of Student as Determinants of Teacher Nonverbal Behavior." *Contemporary Educational Psychology* 4: 324-33.
- Ferguson, Ronald E. 1993. "Spreading the Paradigm of a Master Teacher: The Great Expectations Initiative in Oklahoma," working paper. Taubman Center for State and Local Government, John F. Kennedy School of Government, Harvard University.
- Ford, Donna Y. 1996. *Reversing the Underachievement among Gifted Black Students*. New York: Teacher's College Press.
- Fordham, Signithia, and John Ogbu. 1986. "Black Students' School Success: Coping with the Burden of Acting White." *Urban Review* 18(3): 176-206.
- Gaines, Margie L. 1990. "Accuracy of Teacher Prediction of Elementary Student Achievement." Paper prepared for the annual meeting of the American Educational Research Association.
- Good, Thomas L. 1987. "Two Decades of Research on Teacher Expectations: Findings and Future Directions." *Journal of Teacher Education* 38(4): 32-47.
- Gross, Susan. 1993. "Early Mathematics Performance and Achievement: Results of a Study within a Large Suburban School System." *Journal of Negro Education*. 62: 269-87.
- Guskey, Thomas R. 1982. "The Effects of Change in Instructional Effectiveness on the Relationship of Teacher Expectations and Student Achievement." *Journal of Educational Research* 75: 345-48.
- Haller, Emil J. 1985. "Pupil Race and Elementary School Ability Grouping: Are Teachers Biased Against Black Children?" *American Educational Research Journal* 22(4): 465-83.
- Hanushek, Eric A. 1994. *Making Schools Work: Improving Performance and Controlling Costs*. Brookings.
- Hanushek, Eric A., and Dale W. Jorgenson, eds. 1996. *Improving America's Schools: The Role of Incentives*. Washington: National Academy Press.
- Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.
- Hoge, Robert, and Robert Butcher. 1984. "Analysis of Teacher Judgments of Pupil Achievement Level." *Journal of Educational Psychology* 76: 777-81.
- Irvine, Jacqueline Jordan. 1985. "The Accuracy and Stability of Teachers' Achievement Expectations as Related to Students' Race and Sex." Paper prepared for the annual meeting of the American Educational Research Association.
- _____. 1990. *Black Students and School Failure: Policies, Practices, and Prescriptions*. Greenwood Press.
- Jones, Joseph. 1997. "The Message Project, Fort Wayne Urban League, Phase I." Urban League of Fort Wayne, Indiana, and Taylor University.
- Jussim, Lee. 1989. "Teacher Expectations: Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy." *Journal of Personality and Social Psychology* 57: 469-80.
- Jussim, Lee, and Jacquelynne Eccles. 1992. "Teacher Expectations II: Construction and Reflection of Student Achievement." *Journal of Personality and Social Psychology* 63: 947-61.
- Jussim, Lee, Jacquelynne Eccles, and Stephanie Madon. 1996. "Social Perception, Social Stereotypes, and Teacher Expectations: Accuracy and the Quest for the Powerful Self-Fulfilling Prophecy." *Advances in Experimental Social Psychology* 28: 281-387.
- Katz, M. 1973. "Attitudinal Modernity, Classroom Power and Status Characteristics: An Investigation." Paper prepared for the annual meeting of the American Educational Research Association.
- Kleinfield, Judith. 1972. "The Relative Importance of Teachers and Parents in the Formation of Negro and White Students' Academic Self-Concepts." *Journal of Educational Research* 65: 211-12.
- Klugeel, James R. 1990. "Trends in Whites' Explanations of the Black-White Gap in Socio-economic Status, 1977-1989." *American Sociological Review* 55 (August): 512-25.
- Krupczak, W. P. 1972. "Relationships among Student Self-Concept of Academic Ability, Teacher Perception of Student Academic Ability and Student Achievement." Ph.D. dissertation, University of Miami.
- Ladd, Helen R., ed. 1996. *Holding Schools Accountable: Performance-Based Reform in Education*. Brookings.
- Laasock, E. 1969. *Teaching and Learning in City Schools*. Basic Books.

- Lightfoot, Sara Lawrence. 1978. *Worlds Apart: Relationships between Families and Schools*. Basic Books.
- Merton, Robert. 1948. "The Self-Fulfilling Prophecy." *Antioch Review* 8: 193-210.
- Mickelson, Roslyn A. 1990. "The Attitude-Achievement Paradox among Black Adolescents." *Sociology of Education* 63: 44-61.
- Miller, I. Scott. 1995. *An American Imperative: Accelerating Minority Educational Advancement*. Yale University Press.
- Mitman, Alexis L. 1985. "Teachers' Differential Behavior toward Higher and Lower Achieving Students and its Relation to Selected Teacher Characteristics." *Journal of Educational Psychology* 77: 149-61.
- Monk, Martin J. 1983. "Teacher Expectations? Pupil Responses to Teacher Mediated Classroom Climate." *British Educational Research Journal* 9(2): 153-66.
- Ogbu, John. 1978. *Minority Education and Cast: The American System in Cross-Cultural Comparison*. Academic Press.
- _____. 1983. "Minority Status and Schooling in Plural Societies." *Comparative Education Review* 27(2): 168-203.
- _____. 1987. "Opportunity Structure, Cultural Boundaries, and Literacy." In Judith Langer, ed., *Language, Literacy, and Culture: Issues of Society and Schooling*. Norwood, N.J.: Ablex Press.
- Pedulla, Joseph J., Peter W. Airasian, and George E. Madaus. 1980. "Do Teacher Ratings and Standardized Test Results of Students Yield the Same Information?" *American Educational Research Journal* 17(3): 303-07.
- Puma, Michael, and others. 1993. "Prospectors: The Congressionally Mandated Study of Educational Growth and Opportunity," interim report. Prepared for the U.S. Department of Education, Planning and Evaluation Service.
- Raudenbush, Stephen W. 1984. "Magnitude of Teacher Expectancy Effects on Pupil IQ as a Function of the Credibility of Expectancy Induction: A Synthesis of Findings from 18 Experiments." *Journal of Educational Psychology* 76(1): 85-97.
- Rosenthal, Robert. 1994. "Interpersonal Expectancy Effects: A 30-Year Retrospective." *Current Directions in Psychological Science* 3(6): 176-79.
- Rosenthal, Robert, and I. Jacobson. 1968. *Pygmalion in the Classroom*. Holt, Rinehart, and Winston.
- Rowe, Mary Budd. 1986. "Wait Time: Slowing Down May Be a Way of Speeding Up!" *Journal of Teacher Education* 37(1): 43-50.
- Rubovits, Pamela C., and Martin L. Machir. 1973. "Pygmalion Black and White." *Journal of Personality and Social Psychology* 25: 210-18.
- Shavelson, Richard J., Joel Caddwell, and Tomia Lau. 1977. "Teachers' Sensitivity to the Reliability of Information in Making Pedagogical Decisions." *American Educational Research Journal* 14(Spring): 83-97.
- Slavin, Robert E. 1987. "Mastery Learning Reconsidered." *Review of Educational Research* 57(2): 175-213.
- Smith, Mary Lee. 1980. "Teachers' Expectations." *Evaluation in Education* 4: 53-56.
- Smith, Tom W. 1990. *Ethnic Images*. GSS topical report 19. University of Chicago, National Opinion Research Center (December).
- Snyderman, Mark, and Stanley Rothman. 1986. "Science, Politics, and the IQ Controversy." *Public Interest* 83(Spring): 85.

- _____. 1987. "Survey of Expert Opinion on Intelligence and Aptitude Testing." *American Psychologist* 42(February): 138-39.
- Solorzano, Daniel G. 1992. "An Exploratory Analysis of the Effects of Race, Class, and Gender on Student and Parent Mobility Aspirations." *Journal of Negro Education* 61: 30-44.
- Taylor, Marylee C. 1979. "Race, Sex, and the Expression of Self-Fulfilling Prophecies in a Laboratory Teaching Situation." *Personality and Social Psychology* 6: 897-912.
- Weinstein, Rhonda S. 1985. "Student Mediation of Classroom Expectancy Effects." In Dusek, ed., *Teacher Expectancies*.
- Weinstein, R. S., and others. 1987. "Pygmalion and the Student: Age and Classroom Differences in Children's Awareness of Teacher Expectations." *Child Development* 58: 1079-92.
- Wigfield, Allan, and others. 1991. "Transitions at Early Adolescence: Changes in Children's Domain-Specific Self-Perceptions and General Self-Esteem across the Transition to Junior High School." *Developmental Psychology* 27: 552-65.
- Wills, Sherry. 1972. "Formation of Teachers' Expectations of Students' Academic Performance." Ph.D. Dissertation, University of Texas at Austin.
- Wills, Sherry, and Jerome Brophy. 1974. "The Origins of Teachers' Attitudes towards Young Children." *Journal of Educational Psychology* 66(4): 520-29.
- Yee, Albert H. 1968. "Interpersonal Attitudes of Teacher and Disadvantaged Pupils." *Journal of Human Resources* 3(3): 327-45.