

# TEACHERS' PERCEPTIONS AND EXPECTATIONS AND THE BLACK-WHITE TEST SCORE GAP

RONALD F. FERGUSON  
*Harvard University*

---

---

*Scholars have documented that Black students enter kindergarten with weaker reading skills than their White counterparts and that this disparity sometimes persists through secondary school. This Black-White performance gap is even more evident when comparing students whose parents have equal years of schooling. This article evaluates how schools can positively affect this disparity by examining two potential sources for this difference: teachers and students. It provides evidence for the proposition that teachers' perceptions, expectations, and behaviors interact with students' beliefs, behaviors, and work habits in ways that help to perpetuate the Black-White test score gap.*

**Keywords:** *teacher expectations; teacher perceptions; achievement gap; teacher bias; self-fulfilling prophecy*

---

---

**African American children arrive** at kindergarten with fewer reading skills than Whites, even when their parents have equal years of schooling (Phillips, Crouse, & Ralph, 1998, chapter 5). 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.<sup>1</sup> At worst, the gap widens.<sup>2</sup> Among Blacks and Whites with equal current scores, Blacks tend to make less future progress. This article addresses some of the ways that

---

AUTHOR'S NOTE: *Thanks to Karl Alexander, Bill 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 article. Jason Snipes provided able research assistance.*

URBAN EDUCATION, Vol. 38 No. 4, July 2003 460-507  
DOI: 10.1177/0042085903254970  
© 2003 Corwin Press, Inc.

schools might affect this story. It examines evidence for the proposition that teachers' perceptions, expectations, and behaviors interact with students' beliefs, behaviors, and work habits in ways that help to perpetuate the Black-White test score gap.<sup>3</sup>

No matter what material resources are available, no matter what strategies districts use to allocate children to schools, and no matter how children are grouped for instruction, children spend their days in social interaction with teachers and other students. As students and teachers immerse themselves in the routines of schooling, both perceptions and expectations reflect and determine the goals that both students and teachers 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.

This article examines the controversial but common assumption that teachers' perceptions, expectations, and behaviors are biased by racial stereotypes. The literature is full of seeming contradictions. For example, Sara Lawrence Lightfoot (1978) wrote,

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. (pp. 85-86)

Similarly, Baron, Tom, and Cooper (1985) wrote,

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. (p. 251)

Conversely, doubting that bias is important, Brophy (1985) wrote, "Few teachers can sustain grossly inaccurate expectations for many

of their students in the face of daily feedback that contradicts those expectations (p. 304).” In addition, Haller (1985) observed,

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.(p. 481)<sup>4</sup>

Some aspects of this debate are substantive, but others are semantic. This article begins by distinguishing among alternative definitions of racial bias and reviewing evidence on teachers’ perceptions and expectations. Later sections address ways that teachers’ and students’ behaviors might be both causes and consequences of racially disparate perceptions and expectations regarding achievement and, therefore, contribute to perpetuating the Black-White test score gap.

#### **BIAS IN TEACHERS’ PERCEPTIONS AND EXPECTATIONS**

Expectations, perceptions, and behaviors that look biased if judged by one criterion often look unbiased if judged by another. However, writers on racial bias seldom evaluate their findings by more than a single standard. The discourse that results can be quite perplexing, as one body of literature alleges bias and another denies it. Much of this disagreement is really about what we mean 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 “unconditionally” race neutral. By this criterion, teachers who are unbiased expect the same on average of Blacks and Whites. A second type of benchmark is “conditionally” race neutral—conditioned on observable, measurable criteria. For example, teachers should expect the same of Black and White students on the condition that they have the same past grades and test scores. A third type of benchmark is con-

ditioned not on past performance but instead on unobserved potential. It requires neutrality among—for example, equal expectations and aspirations for—Blacks and Whites who have equal “potential.” Unfortunately, insofar as “potential” differs from past performance, it is difficult to prove. Assuming that Black and White children have the same potential at birth, which seems a fair assumption, then there is no distinction at birth between unconditional race neutrality and neutrality conditioned on unobserved potential. However, as children grow older, disparities in potential may develop if experience alters potential (e.g., consider recent literature on brain development). Thus, as children grow older, unconditional race neutrality may or may not remain the best approximation to neutrality conditioned on unobserved potential.

#### UNCONDITIONAL RACIAL NEUTRALITY

Unconditional race neutrality requires that perceptions, expectations, and behaviors be uncorrelated with 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. The typical finding in such studies is that teachers are racially biased.

Researchers in experimental studies fabricate information about students. The information includes race in addition to other characteristics, but the sample is selected to avoid any correlation of these other characteristics with race.<sup>5</sup> In a typical experiment, teachers receive information about students via written descriptions, photographs, videotapes, or occasionally real children who act as the experimenter's confederates. Teachers then predict one or more measures of ability or academic performance for each student. If the experiment is run well, teachers do not discern that race is a variable in the experiment or that the real purpose is to assess their racial biases.

Baron et al. (1985) conducted a meta-analysis of experimental studies that focused on teachers' expectations, 16 of which dealt with race. Teachers had higher expectations for White students in 9 of the studies and for Blacks in 1 of the studies. Six studies in which the differences were statistically insignificant did not report which

group was favored. Of the 5 studies with statistically significant differences, all favored Whites. In the meta-analysis, the hypothesis of identical expectations for Black and White students is clearly rejected ( $p < .002$ ).<sup>6</sup>

The meta-analysis of Baron et al. (1985) missed one interesting study by DeMeis and Turner (1978) that makes no reference to teachers' expectations, attitudes, or biases in its title.<sup>7</sup> Nevertheless, it supports the same conclusion. In the DeMeis and Turner study, the participants were 68 White, female, elementary schoolteachers drawn from summer school classes at a university in Kentucky during the 1970s. Their teaching experience averaged 7 years. The students in the study were a pool of fifth-grade males responding to the question, "What happened on your favorite TV show the last time you watched it?" Each tape was accompanied by a picture of a Black or White student. DeMeis 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 ( $p < .0001$ ) for each of the four outcomes.<sup>8</sup>

If the benchmark is unconditional racial neutrality, teachers hold racially biased expectations. What should we make of this pervasive racial bias? Consider people who learn from real life that the odds of getting heads are 60:40 when flipping coins. Place these people in an experimental situation where, unknown to them, the odds have been set at 50:50. If each person is given only one toss of the coin in the experimental situation, what will he or she predict? Will their predictions be unbiased? If, as in the experiments discussed above, the benchmark for declaring expectations unbiased is unconditional racial neutrality, then biased expectations are what one should expect in an environment where real differences in performance between Blacks and Whites are the norm.<sup>9</sup> For the same reasons, bias of this type is also pervasive in naturalistic studies (i.e., 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 be 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 OF ACCURACY

For more than two decades, scholars in education have emphasized that teachers' contemporaneous perceptions of students' performance as well as their expectations for students' future performance are generally accurate (e.g., see Egan & Archer, 1985; Good, 1987; Hoge & Butcher, 1984; Mitman, 1985; Monk, 1983; Pedulla, Airasian, & Madaus, 1980). For example, first-grade teachers can learn enough about children in the first few weeks of school to predict with some accuracy the students' rank order on exams held at the beginning of second grade (Brophy & Good, 1974, Table 6.1). Once set, teachers' expectations do not change a great deal. This may be because even early impressions of students' proficiencies are accurate, and the actual rank order does not change much.

There could be several reasons for stability in rank orderings. First, teachers' perceptions and expectations might be relatively inflexible. Self-fulfilling expectation effects, discussed below, will typically be strongest for teachers whose expectations are the least flexible.<sup>10</sup> For these teachers, correlations between beginning-of-year and end-of-year assessments should be among the highest.<sup>11</sup> A second reason for stability of class-rank orderings might be that few students try hard to change their positions. A third reason might be that the pace and style of standard teaching offer few effective opportunities for students who are behind to catch up.<sup>12</sup> Most evidence about the accuracy of teachers' perceptions comes from correlations between teachers' predictions and actual test scores, which typically range between .50 and .90 (Brophy & Good, 1974; Egan & Archer, 1985; Evertson, Brophy, & Good, 1972; Irvine, 1985; Willis, 1972). At least in the low end of this range, one could also focus on the inaccuracy of the predictions in "glass half empty" fashion.

I found only three studies that reported separate correlations for Blacks and Whites. Haller (1985) found that teachers' subjective

assessments of fourth, fifth, and sixth graders' reading proficiency correlated .73 with the Comprehensive Test of Basic Skills for White students and .74 for Blacks.<sup>13</sup> Irvine (1990) had teachers rank 213 fifth, sixth, and seventh graders on general academic ability during the 2nd, 10th, and final weeks of the 1983 to 1984 school year.<sup>14</sup> Correlations between these ratings and scores on the California Achievement Test (CAT) were similar for Blacks and Whites.<sup>15</sup> Gaines (1990) also found that teachers' predictions of students' performances on the Iowa Test of Basic Skills were as accurate for Blacks as for Whites.

This 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.<sup>16</sup> Neither does it mean that teachers are racially unbiased. Here, accuracy is not always the opposite of bias. If self-fulfilling prophecy always worked perfectly, for example, each student's performance would be exactly what the teacher expected. If expectations were biased, outcomes would be perfectly predicted but biased.

#### **RACIAL NEUTRALITY CONDITIONED ON OBSERVABLES**

Racial neutrality conditioned on observables is a second type of benchmark for measuring bias. Instead of unconditional racial neutrality, as discussed above, the benchmark here assumes that a teacher's perceptions or expectations are unbiased if they are based only on legitimate observable predictors of performance, such as the student's past grades, test scores, attitudes about school, and beliefs about personal ability. Most people would agree that these are reasonable and "legitimate" things on which to base a prediction of performance. The benchmark in this case is only conditionally race neutral: If past performance is correlated with race, the benchmark will be too. Bias is the difference between the actual perception or expectation and the benchmark.

This type of bias can be identified by regressing the teacher's perception or expectation on both race and the other explanatory

variables that we regard as “legitimate” determinants of teachers’ perceptions or expectations. The coefficient of student 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 are biased in the judgments that they use in nominating students for particular curriculum tracks or ability groups. As we shall see, it might also be more appropriate for use in analyses of whether teachers’ biases produce self-fulfilling prophecies of poor performance for Black students. However, it is not sufficient to distinguish conditional from unconditional neutrality; the existing literature often makes a further distinction between past performance and future potential.

#### RACIAL NEUTRALITY CONDITIONED ON POTENTIAL

A third type of benchmark that may or may not equate with either of the two discussed above, is the level of performance that the student could reach at his or her full potential. Here, the bias is in perception or estimation of a person’s full potential. The full-potential benchmark equals demonstrated plus latent potential. The alleged racial bias is that people underestimate latent potential more for Blacks than for Whites.

A major concern of African Americans is that teachers underestimate Black students’ potential, not necessarily their performance. As an illustration of this concern, consider the following passage from a report by the Committee on Policy for Racial Justice, titled *Visions of a Better Way: A Black Appraisal of American Public Schooling*, published by the Joint Center for Political and Economic Studies in 1989 (as cited in Miller, 1995):

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. (p. 203)<sup>17</sup>

The passage that follows clearly alleges bias, judged by a “future potential” benchmark:

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. (from Leacock as cited in Brophy & Good, 1974, p. 10)

If perceptions of children’s intellectual potential affect goal setting in both homes and classrooms, which surely they must, then teachers and parents who underestimate children’s potential will tend to set goals that are too low.<sup>18</sup> Underestimation of potential is undoubtedly a major problem, irrespective of race. It is a major waste of human potential and a social injustice that we do not give teachers the incentives and supports they need to set, believe in, and skillfully pursue higher goals for all students, but especially for African Americans and other stigmatized minorities. Because we underestimate potential, the payoff to searching more aggressively for ways of helping children would surely be higher than most people imagine.

Bias related to future potential is impossible to estimate reliably because there is no clear basis on which to estimate human potential.<sup>19</sup> Surveys find that expressed beliefs in the intellectual inferiority of Blacks have moderated over the years.<sup>20</sup> For example, the percentage of Whites responding to the General Social Survey that Blacks have less “in-born ability to learn” fell from 27% in 1977 to 10% in 1996 (Jencks & Phillips, introduction; Kluegel, 1990, pp. 514-515, 517). There is no way to know the degree to which this reduction is due to changes in beliefs or changes in social norms. For example, Tom Smith (as cited in Miller, 1995) discussed how the same General Social Survey found that when respondents were not constrained to attribute differences to genetic factors, 53%

agreed that Blacks and Hispanics are less intelligent than Whites; further, 30% of Blacks and 35% of Hispanics agreed (p. 183).<sup>21</sup>

Many experts also think that genetic differences are at least partially to blame for existing Black-White differences in academic achievement. Snyderman and Rothman (as cited in Miller, 1995) discussed a 1984 survey that questioned 1,020 experts on intelligence, most of whom were professors and university-based researchers who study testing, psychology, and education. As Miller (1995) reported, almost half (46%) expressed the opinion that Black-White differences in intelligence are at least partially genetic. Of the others, 15% said that only environment was responsible, 24% regarded the available evidence as insufficient, and 14 did not answer the question (Miller, 1995, pp. 186-187). In other words, only 15% clearly disagreed. With expert opinion slanted so strongly in favor of the genetic hypothesis and widespread media attention to books such as *The Bell Curve*, there is little prospect that "rumors of inferiority" will cease or that racial differences in estimated potential will disappear.<sup>22</sup>

Finally, writers concerned with bias in estimating 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, a more appropriate terminology than "self-fulfilling" might be to say that bias of this type produces expectations that are "sustaining" of past trends.<sup>23</sup> A sustaining 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 arrived.

#### **TYPE II BIAS AND SELF-FULFILLING PROPHECY**

A self-fulfilling prophecy is one that makes a bias in a teacher's expectation regarding a student's performance affect the student's performance. Self-fulfilling prophecies can be associated with any of the three conceptions of bias discussed above, but only those associated with the second type (where the benchmark is condi-

tioned on observables) can be well measured. Although the basic idea was introduced into social science by Merton (1948), Rosenthal and Jacobson's (1968) work sparked a small industry of such studies during the 1970s and early 1980s. The effect shows up (and also fails to show up) in a wide range of experimental studies for both animal and human subjects (Rosenthal, 1994). Experimental studies in education typically involve random assignment of students to groups that have been labeled as high or low performing.

Successful instigation of self-fulfilling prophecies by researchers requires that (a) teachers believe false information given to them about students, (b) teachers act on the information in ways that students can perceive, and (c) students respond in ways that confirm the expectation. The effect can fail to appear, which it often does, if any of these conditions fail.<sup>24</sup>

In experiments that confirm the effect, groups labeled as high performing outperform those labeled as low performing. A meta-analysis by Smith (1980) identified 44 effect-size estimates for reading scores, with an average effect size of 0.48 standard deviations, distinguishing students with high and low labels.<sup>25</sup> The average across 17 effects for math, reported in the same paper, was much smaller at 0.18. Why effects should be smaller for math than for reading is unclear. Perhaps math instruction is less flexible and therefore less affected by teacher perceptions.

Brophy (1985), a leader in the field since the early 1970s, asserted that on average, teacher expectations in real classrooms probably make only a small difference in student achievement (p. 304).<sup>26</sup> He added the caveat, however, that teachers who hold rigid expectations and permit these to guide interactions with their students can produce effects that are larger. It is plausible, but not established in any literature that I have seen, that expectations might be more rigid regarding Black students than 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 with the passage of 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 equation where the dependent variable is the student's actual performance at the end of a school year.<sup>27</sup> 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. When teacher biases exist but do not affect actual scores or grades, it can be 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 his or her expectation for future performance can differ with one showing bias and the other not.<sup>28</sup>

#### TESTING FOR RACIAL DIFFERENCES IN EXPECTANCY EFFECTS

Jussim, Eccles, and Madon (1996) are the only researchers who have tested for racial differences in impacts of teachers' perceptions on test scores (pp. 350-351).<sup>29</sup> Teachers' perceptions of 1,664 sixth graders' current performance, talent, and effort in math were collected in October of the 1982 to 1983 school year.<sup>30</sup> Jussim et al. then tested for what they called "racial stereotype bias" (whether student race predicts teachers' October perceptions after controlling for previous grades, previous test scores, self-concept of math ability, self-reported level of effort, and self-reported time spent on homework). This type of bias uses the second type of benchmark defined above—racial neutrality conditioned on observables, including past performance. Jussim et al. found no evidence of racial stereotype bias in teachers' perceptions of *current* performance, talent, or effort for this sample of sixth graders.<sup>31</sup> The coefficient on student race was small and statistically insignificant.<sup>32</sup>

If past performance and attitudes explain racial differences in teachers' *current* perceptions, then these perceptions can be an important source of the *future* Black-White test score gap only if teachers' perceptions affect Blacks and Whites differently. This is precisely what Jussim et al. (1996) found. They analyzed the effects of teachers' perceptions of performance, talent, and effort in Octo-

ber on both math grades and scores on the math section of the Michigan Educational Assessment Program for May of the spring semester of the 1982 to 1983 school year.<sup>33</sup>

For both grades and scores, Jussim et al. (1996) found that the estimated impact of teacher perceptions was almost 3 times as great for African Americans as for Whites.<sup>34</sup> Effects were also larger for girls and for children from low-income families. Further, the effect was cumulative across disadvantages or stigmas: Black children from low-income backgrounds experienced the effects of both race and income. Teachers' perceptions of student effort do not affect Michigan Educational Assessment Program scores, but they do affect grades, even though they are not strongly correlated with students' self-reports of effort.<sup>35</sup>

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.<sup>36</sup> A more likely explanation is that teachers are less flexible in their expectations for Blacks, females, and students from low-income households. Or as Weinstein (1985) speculated,

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. (p. 344)

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. Two sections below address these possibilities.

Table 1 shows simulated math scores and grades (in standard deviation units) cross-tabulated with teachers' performance ratings for students, holding all other student characteristics constant. For both Blacks and Whites, there is a positive relationship between the teacher's October performance rating and the student's grades and scores in May. However, the effect is stronger for Blacks. Blacks who receive performance ratings in the top category (e.g., Category 5)

**TABLE 1**  
**Spring Standardized Grades and Test Scores in Mathematics**  
**and Fall Performance Ratings, Sixth Grade, 1982-83<sup>a</sup>**

<i>Measure and Race</i>	<i>Fall Performance Rating</i>				
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
Predicted spring grades					
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
Predicted spring scores					
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.

in October are predicted to outperform Whites who receive the same rating (again, assuming equal values for control variables). Conversely, Blacks who receive the lowest October performance ratings lag an estimated half standard deviation behind Whites who got the same ratings.

If teachers tend to be accurate in their current perceptions and in their expectations for future progress, then the findings of Jussim et al. (1996) require that teachers should expect the pattern shown in Table 1.<sup>37</sup> This would be 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 a self-fulfilling prophecy of 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. The next two sections address these possibilities.

### **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. A recent study by Casteel (1997) asked eighth and ninth graders whom they want most to please with their class work.<sup>38</sup> "Teachers" was the answer for 81% of Black females, 62% of Black males, 28% of White females, and 32% of White males. Whites were more concerned with pleasing parents. Entwisle and Alexander (1988) found that teachers' ratings of maturity for first graders have larger effects for Blacks than for Whites on both verbal and arithmetic scores on the CAT. Kleinfeld (1972) found that high school students' concept of their ability is more correlated with perceived teacher ratings of ability for Blacks but more correlated with perceived parent ratings for Whites. Irvine (1990) reached similar conclusions.<sup>39</sup>

Jussim et al. (1996) suggested, and I agree, that Claude Steele's recent work (e.g., see Steele and Aronson in this volume) offers one reason why Black and White students might respond differently faced with identical classroom conditions.<sup>40</sup> What Steele called "stereotype threat" and the resulting "stereotype anxiety" can operate for members of any stigmatized group. When the stereotype concerns ability, individuals fear performing in ways that might corroborate the stereotype. They fear that the stereotype might become the basis of others' pejorative judgments, as well as their own self-perceptions.

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

Steele tested this idea only at the college level, and even there, his findings were only for high-achieving students at highly selec-

tive colleges. The degree to which stereotype threat and anxiety might apply for students in primary and secondary schools remains to be investigated. The findings of Jussim et al. (1996) were for sixth graders. Are children this young aware enough of stereotypes to be susceptible to stereotype threat or anxiety?

Perhaps.<sup>41</sup> Gross (1993) studied math performance for students in a racially integrated suburb of Washington, D.C., during the 1985 to 1986 school year. In the fourth grade, 92% of Blacks and 86% of Whites who were above grade level on the number of math competencies that they had mastered scored in the 8th and 9th stanines (i.e., ninths) of the CAT. By sixth grade, 82% of Whites who were above grade level in completion of competencies were still in the 8th and 9th stanines on the CAT. For Blacks, however, the figure was only 68%.<sup>42</sup> Gross pointed out that this pattern of performance on the CAT for sixth graders was inconsistent with students' in-school performance, and she used it to caution against basing ability-group placements only on test scores.

Gross (1993) and her team also conducted focus groups with middle and high school students. She reported "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" (p. 283). 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. (p. 283)

Although the focus groups were for middle and high school students, Gross (personal communication, February 1998) reported that it was common knowledge that children in upper elementary grades felt the same types of pressures. If the CAT for sixth graders was regarded as a test of ability, which it may well have been, then Steele's (1998) theory could be an important reason why Black sixth graders who were above grade level on competencies got lower CAT scores than White peers with the same level of classroom competence.<sup>43</sup>

Gross (1993) also reported what appears to be the type of disengagement that Steele hypothesized—a type that could help explain the larger negative impact on Black students that Jussim et al. (1996) found when performance was perceived to be low. Gross reported that teachers said Black students were overrepresented among students who were least studious and least well prepared for class: They “did not come to their classes prepared to work or in the proper frame of mind to attend fully to instruction.” (1993, p. 281)

In addition, both teachers and administrators reported that Black parents were less supportive of the school’s mission than White parents. Meanwhile, Black parents, when convened in focus groups, were the most supportive of the idea that their children should strive for the higher level math classes, even if that meant lower grades. White parents were more prone to say that their children should stay in the top sections only if they were likely to do well. Were Black parents sending the mixed message, “Shoot for the top, but if you don’t do as well as the White kids, we’ll understand”? Could this have contributed to the ambivalence that their children’s work habits appeared to express? (See Figure 1 below regarding racial comparisons of work habits and other attitudes.) If Black children sense more ambivalence from their parents than White children do, then teachers’ opinions might take on a special significance, as the statistical evidence discussed above appears to show.

In a study inspired by the work of John Ogbu (1978, 1983, 1987; Fordham & Ogbu, 1986), Mickelson (1990) distinguished “abstract” from “concrete” attitudes. Mickelson found that concrete attitudes predict cumulative high school grade point average but abstract attitudes do not.<sup>44</sup> Her measure of abstract attitudes reflects mainstream ideology—standard optimistic rhetoric about education and the American dream. In contrast, her measure of concrete attitudes includes questions that illicit doubt and ambivalence about education as a route to success in mainstream society.<sup>45</sup> Students might acquire concrete attitudes from routine, informal, personal interaction with friends, parents, other adults, and the broader society. Mickelson found that Blacks agreed to an even greater degree than Whites with the optimistic but abstract beliefs about success and the American dream. However, Blacks’ concrete

**TABLE 2**  
**Mean Scores on Measures of "Abstract" and "Concrete"**  
**Beliefs About the Importance of Education for Success,**  
**by Race and Class Background for 12th Graders in**  
**Eight Los Angeles High Schools, Spring 1983**

	<i>Black Male</i>	<i>White Male</i>	<i>Effect Size</i>	<i>Black Female</i>	<i>White Female</i>	<i>Effect Size</i>
For children of white -collar parents						
Abstract score	5.50	5.06	0.58	5.27	5.09	0.24
Concrete score	4.38	4.90	-0.53	4.43	5.00	-0.58
Sample size	56	224		84	241	
For children of blue- collar parents						
Abstract score	5.28	4.99	0.38	5.34	5.21	0.17
Concrete score	4.19	4.54	-0.36	4.19	4.81	-0.63
Sample size	138	100		140	93	

SOURCE: Mickelson, 1990.

NOTE: Full sample standard deviations: abstract scores ( $SD = .76$ ); concrete scores ( $SD = .98$ ). Racial differences in abstract scores,  $p < .05$ ; class differences in concrete scores,  $p < .0005$ ; racial differences in concrete scores,  $p < .0001$ .

attitudes were less hopeful than Whites'. Table 2 summarizes the pattern.<sup>46</sup> It suggests a possible reason why surveys usually find that Blacks subscribe to mainstream values as much as Whites do but then behave in ways that show less commitment to mainstream success.

### DO TEACHERS TREAT BLACK AND WHITE STUDENTS DIFFERENTLY?

I found only four experimental studies dealing with teachers' treatment of Black and White students. All date from the 1970s and early 1980s (Coates, 1972; Feldman & Orchowsky, 1979; Rubovits & Maehr, 1973; Taylor, 1979).<sup>47</sup> These studies control differences between students (or "virtual" or "phantom" 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 found that teachers were less supportive of Black than White students.

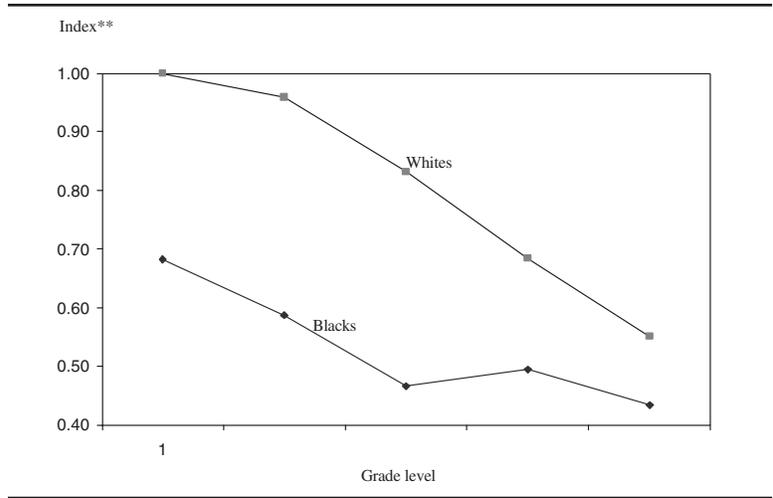
Taylor (1979) conducted an experiment in which a 6-year-old student was said to be watching from behind a screen as the participants, college students in teacher training, taught a prescribed lesson.<sup>48</sup> Taylor found that when these phantom students were Black, they received briefer feedback after mistakes (standardized effect size = 0.613), less positive feedback after correct responses (effect size = 0.423), and fewer “helpful slips of the tongue” (i.e., unauthorized coaching) (effect size = 0.536). Each of the experimental studies suggests that some teachers may be helping Whites more than Blacks and that the differences may be large enough to have nontrivial effects on performance.

Studies of real classrooms confirm this hypothesis. These studies sometimes find no racial differences but more frequently have found differences favoring Whites.<sup>49</sup> Of course, studies finding differences are probably more likely to be published. Nonetheless, if our benchmark is unconditional racial 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 we controlled racial differences in students’ behaviors, work habits, and social skills. But because students and parents cannot read a teacher’s mind, they may *think* that differences in treatment reflect unjustified and unfair favoritism. When teachers appear biased, trust may be eroded and relationships spoiled (see below).

#### EVIDENCE ON POSSIBLE REASONS FOR DIFFERENTIAL TREATMENT

Typically, we have no way of knowing whether teachers’ perceptions of students’ attitudes and behaviors are accurate. The finding of Jussim et al. (1996) that teacher perceptions of effort predicted grades but not standardized test scores calls attention to the possibility that perceptions are inaccurate; so does the finding of Jussim et al. that teachers’ perceptions of effort were only moderately correlated with students’ own self-reports.

In 1990, the *Prospects* survey asked a national sample of teachers to rate specific students in their classes on, “Cares about doing well,” “Gets along with teachers,” and “Works hard at school” (for



**FIGURE 1 Teachers' Perceptions of Students' Levels of Effort\***

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).

\* 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.

\*\* 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.

more on the *Prospects* survey, see Puma, Jones, Rock, & Fernandez, 1993). The response categories ranged from “very much” to “not at all.” Teachers rated Black children lower than Whites on all three items.<sup>50</sup> To measure racial differences in these ratings, I subtracted for each racial group, the percentage with the lowest ranking (i.e., “not at all”) from the percentage with the highest ranking (i.e., “very much”) in each grade, and then summed 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. Figure 1 shows the means for Blacks and Whites by grade level.

Figure 1 shows that teachers perceived the greatest Black-White difference in the early elementary years. After the fifth grade, the

gap narrows but does not completely close. The similarity in how Blacks and Whites relate to school by ninth grade is consistent with data on eighth graders from the National Longitudinal Survey (NELS88). Several authors, including Cook and Ludwig in this volume, have remarked on the similarity of Black and White attitudes in NELS88 (see also Miller, 1995; Solorzano, 1992). Because NELS88 did not reach students earlier in their school careers (and also because it does not ask the right questions), it may lead people to underestimate the degree to which Black-White differences in work habits, behavior, and attitudes prior to eighth grade affect teacher-student relations and the Black-White test score gap.

#### RECIPROCAL EFFECTS OF STUDENT BEHAVIOR AND TEACHER PERFORMANCE

Teachers' judgments about how much they enjoy teaching students inevitably affect teachers' behaviors. This can apply to entire classrooms and to individual students. Teachers may respond to difficult students by withdrawing support (Brophy & Good, 1974).<sup>51</sup> This could help account for the finding of Jussim et al. (1996) that teachers' beginning-of-year perceptions, although initially unbiased, are stronger predictors of end-of-year performance for Blacks. Specifically, between equally low-performing Black and White students, Black students might be perceived as more difficult (and therefore receive less teacher support than Whites), and between equally high-performing Black and White students, Black students might be perceived as less difficult (and therefore receive more teacher support than Whites).<sup>52</sup> I found no quantitative research in naturalistic settings that controlled for initial student performance and then systematically measured racial differences in how much positive reinforcement teachers provided. However, Black *students* on average may give *teachers* less positive reinforcement than White students with the similar beginning-of-year performance.

For example, Willis and Brophy (1974) asked 28 first-grade teachers to nominate three children to each of four groups. The groups, along with Willis and Brophy's summary of teachers' responses, were as follows:

*Attachment group:* 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 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 group:* 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 group:* 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 were] perceived as making legitimate demands because they generally conform to classroom rules but are in need of help due to low ability.

*Rejection group:* 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 nonWhite 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. (p. 132)

The rejection group was the only one in which nonWhite boys were overrepresented 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. We know from Figure 1 that teachers in the 1990s perceived 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 teachers' and students' races 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 can present.<sup>53</sup> The following is a quote from a young Black teacher, Paula:

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 she attended] 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. (Cabello & Burstein, 1995, pp. 289-290)

It remains an open question how much difference such feelings make for children's outcomes. Many "burned out" teachers in difficult schools may simply be going through the motions, waiting for retirement. Also, it is unclear to what degree this pattern is racially distinct. In many classrooms, teachers and students are embroiled in conflicts and confusion that they lack the skills and external supports to resolve. Paula was enrolled in a program to help teachers understand, manage, and teach difficult students. Research analyzing the effectiveness and replicability of such programs should be a priority if we are to improve the schooling of Black children in settings where behavior is a problem.

Teachers also need to communicate more clearly about the quality of children's performance in classrooms where achievement correlates closely with race. Signals about performance have racial overtones in such classrooms. This can interfere with teacher-student

relations and learning. A summary of focus group discussions with 9- to 19-year-old Black males in a Midwestern city 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 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. (Jones, 1997, p. 9)

Teachers in integrated schools can be “biased” in ways as simple as reinforcing a propensity of White children to speak more often in class (e.g., see Katz as cited in Brophy & Good, 1974; Irvine, 1990). Black students may assume that this means teachers think Whites are “smarter” or like the White students more. Ways that 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 stay engaged and aim for mastery.<sup>54</sup>

### **RESPONSIVE TEACHING**

The average Black child arrives at kindergarten with fewer academic skills than the average White child. Schools may then simply push children along in ways that sustain or add to racial disparities, validating the expectation that Black-White differences in achievement are normal or perhaps even inevitable. But if instruction is appropriately stimulating and responsive to children's progress, teachers' expectations may not be either self-fulfilling or sustaining. The more inviting and responsive instruction is to children's 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” or how long a teacher waits for students to raise their hands, to begin talking, and 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 (Rowe, 1986).<sup>56</sup>

Corrective feedback is probably more important than wait time.<sup>57</sup> In a study where race was not mentioned, Guskey (1982) studied 44 intermediate and high school teachers who taught various subjects in two metropolitan school systems. Each teacher taught two matched classes after receiving training in a particular approach to mastery learning. The teacher taught one class using his or her standard methods and another using a "feedback and corrective" process learned in the training.<sup>58</sup> Both classes used the same final exam and grading standards. Guskey compared teacher ratings of "probable achievement" from early in the semester with final grades and exam scores and with ratings of "achievement potential" that teachers gave at the end of the term. For 10 teachers, the training made no difference to their students' performance. However, for 34 other teachers (called the "positive change" group) the experimental classes did better than the control classes on both grades and scores. Among these 34, the improved techniques for feedback and correction made teachers' early expectations less predictive of later achievement. Specifically, correlations between teachers' initial ratings of probable achievement and the students' final grades and exam scores were markedly lower for experimental classes than for classes using customary methods; teachers in the "no-change" group had high correlations in both classes (see Table 3). It seems likely that better feedback and corrective methods could also affect the rank order of performance by race, but Guskey did not investigate this issue.

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. Brattesani, Weinstein, and Marshall (1984), in an often-cited study, compared fourth-, fifth-, and sixth-grade classes where student surveys showed high and low levels of differential treatment.<sup>59</sup> In classrooms with more differential 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.<sup>60</sup>

**TABLE 3**  
**Median Correlations Between Teachers' Initial Ratings of**  
**Students, Final Ratings, Students' Course Grades,**  
**and Students' Course Examination Scores**

	<i>Correlation Between Initial Rating and</i>					
	<i>Final Rating</i>		<i>Course Grade</i>		<i>Final Exam Score</i>	
	<i>Experi- mental</i>	<i>Control</i>	<i>Experi- mental</i>	<i>Control</i>	<i>Experi- mental</i>	<i>Control</i>
Positive change ( <i>N</i> = 34)	.53	.83	.51	.80	.31	.50
No change ( <i>N</i> = 10)	.92	.90	.77	.79	.69	.75

SOURCE: Guskey, 1982.

NOTE: Of the 44 teachers in the sample, 34 were labeled the "positive change" group because their experimental classes had higher final exam scores and course grades than their control classes. For the 10 teachers in the "no change" group, either grades or scores on the final exam were higher in the control class.

Both Guskey (1982) and Brattesani et al. (1984) showed that greater responsiveness to individual children can weaken the link between past and future performance, perhaps also altering trajectories. Both were silent about race and ethnicity. However, studies that deal directly with race do not investigate whether teaching practices can change the rank order of performance among students.

### THE GREAT EXPECTATIONS (GE) INITIATIVE

The Great Expectations (GE) initiative is a public-private partnership created in 1989 to bring Marva Collins's ideas about teaching into Oklahoma schools.<sup>61</sup> By ending with the GE story, my purpose is to show real people struggling, with some success, to change teaching practices and, in the process, teachers' expectations for disadvantaged, mostly minority children. GE includes a range of techniques that Collins developed. It aims to nurture in *every* student, 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. Misbehaving students should be

reminded regularly that the teacher cares and refuses to give up on them. Methods combine high challenge for students with forms of feedback from teachers *and peers* that make learning fun and emphasize its importance for happy and effective lives. 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 that students memorize to discuss and recite at school and at home.

The incentive for schools to participate in the GE initiative was a threat of takeover by the state if test scores for third graders persisted below the 25th percentile for 3 consecutive years on the Iowa Test of Basic Skills. Educators in schools that joined GE had heard of Collins. Many had seen her on CBS's *60 Minutes* and knew of her reputation for working wonders with children in inner-city Chicago. Collins's school had never been independently evaluated, but she appeared to be effective with the types of children that schools in Oklahoma were failing. Administrators were not certain that Collins's methods were transferable from Chicago to Oklahoma, but they judged it worth a try. The first training took place at Collins's Westside Preparatory School in Chicago for two teachers from each of 25 pilot schools. In Chicago, they had a "seeing is believing" experience concerning what children from disadvantaged backgrounds could achieve. Thus confronted, they became students of Collins' methods.

Nevertheless, as the initiative spread through the 25 schools, there was substantial resistance from other teachers who had not gone to Chicago. The resistance came in three basic forms:

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."

The head mentor teacher told me that virtually all of the resistance she encountered combined one or more of these three perspectives.

Her responses to the excuses were sometimes successful, sometimes not.<sup>62</sup>

Some teachers were insecure. Their low expectations for students were partly the consequence of low expectations for themselves as teachers.<sup>63</sup> At the other extreme were people who thrived using Collins's ideas and felt professionally rejuvenated. A baseline survey that I administered at a summer institute in July 1993, asked respondents who had prior GE training to describe the changes in their classrooms since they began using GE methods. The questions were worded as follows: "Because of Great Expectations, the improvement in the [aspect of classroom performance] of my students has been: 1 = *More than I thought was possible*, 2 = *A lot*, 3 = *Some*, 4 = *None*." The percentage giving each answer for each of the four aspects of classroom performance is shown in Table 4.

The following two teachers are "existence proofs" of the proposition that teaching practices and expectations can change dramatically, even among experienced teachers. At the same time, both teachers expressed reservations.<sup>64</sup>

**GREG ROBARTS, FOURTH-GRADE TEACHER,  
BEEHIVE ELEMENTARY SCHOOL**

Robarts's classroom is an almost even mix of Black, White, Chicano, and Native American children, almost all of whom are from very poor families. Before GE, Robarts had taught for 17 years and believed himself to be a good teacher. But seeing what children at Westside Preparatory School in Chicago could do gave him pause. He said, "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 17 years teaching." He described 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.

**TABLE 4**  
**Teachers' Assessments of Student Progress Due to the**  
**Great Expectations Program in Oklahoma (N = 76)**

<i>Teachers' Assessment</i>	<i>Academic Performance</i>	<i>Attitudes</i>	<i>Behaviors</i>	<i>Teachers' Job Satisfaction</i>
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: Tabulations from a survey that I conducted at the Summer Institute for Great Expectations in July 1993. The summer institute was targeted primarily at teachers who had no prior training. This sample of 76 teachers represents close to a 100% response rate from the previously trained teachers who attended during the week that the survey was administered.

Collins's basic philosophy resonated with Robarts's beliefs about teaching, but he had to unlearn old habits: sitting at his desk and saying, "Open the book to page 34, here are the instructions. . . ." Even his own principal described Robarts as a "desk sitter" before he changed to the GE way of running his classroom: "Teach on your feet, not in your seat." Before GE,

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.

[Question: 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.

People only vaguely familiar with the GE approach often say it overemphasizes memorization and underemphasizes higher order thinking. Robarts disagreed. He said memorization and higher order thinking are complements, not substitutes. Memory is the foundation for higher order thinking. He found that many children cannot remember a dictated sentence long enough to write it on their papers. After a few weeks of memorizing poetry and other things, he said, the change is remarkable. He thinks people who dismiss memory work as outmoded are simply uninformed concerning how children learn, not only because memory supports higher order thinking, but also because children can memorize things that are worth knowing. In addition, recitation by children of what they have memorized builds self-confidence and motivation. Robarts said,

If you had told me 2 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 GE classrooms is memorization of poetry that contains rules for healthy and productive living—messages worth remembering.

Robarts reported exciting results from his efforts: "Absenteeism is almost nil, refusal to do homework is almost nil, test scores are substantially up." Also, he had a "miracle turn-around student" during his first semester using Collins's methods. The student's disciplinary folder was a "blizzard of suspensions." He was diagnosed as learning disabled, Robarts said, a "throw-away child":

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 next year, his classmates teased him for speaking standard English. Jerry persisted, with support from his new teacher. On returning to visit Robarts, Jerry reported that his new classmates began to change to be more like him because of the positive responses he was getting from the teacher.

Robarts realizes that he is not typical. Other teachers need more ongoing support. I first interviewed Robarts in December of 1991. This was after the first summer institute but 9 months before mentor teachers began working in teachers' classrooms. At that time, Robarts was sober in his comments about the value of a 4-hour demonstration that he was scheduled to give the next week, with his class, for teachers who had not gone to the previous 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.

Robarts was among the teachers who pushed for establishment of the summer institutes and the use of mentor teachers for ongoing technical assistance. Both were instituted and remain in place, complemented now by an academy for principals that has been in place for the past few years.

**MRS. GLORIA CHAVERS, THIRD-GRADE TEACHER,  
LAFAYETTE ELEMENTARY SCHOOL**

Chavers recalled not thinking much about GE at the time that her school applied to participate. However, she noted that when the first two teachers came back from Chicago, "They were excited. There

was no doubt about that.” The principal asked other teachers to visit the classrooms of the trained teachers to observe.

So we went to Mrs. Sherrin’s 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 5-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 recalled that she saw changes that she did not like in Sherrin’s students. As a second-grade teacher, she had taught some of the same children whom Sherrin was teaching in the third grade. Now, the children were calling themselves “college bound.” Chavers said her view at the time 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. Chavers said, “As it turned out, it was really well worth it.” The next semester, Chavers 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 reported 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.” Chavers learned how to teach addition and subtraction using chants at the summer institute, but the class at the summer institute had not gotten to multiplication and division. So, Chavers made up her own chants for multiplication and division. She recalled,

And then I told them [her 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 cannot 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 talked on about bringing her love of music into the classroom now in ways she never felt authorized to do before. She talked about impressing her friends with her students' written work. She spoke 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 they themselves did not see until junior high school. Parental interest and participation has clearly increased. According to the district superintendent, "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 it made for their children.

Chavers said that her children are convinced that they can do anything. When she plays choral music on the 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 2 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 where she got the permission, she said, "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 about new freedom to bring their personal interests and talents into their classrooms.

At the time of this interview, Chavers had been teaching for 17 years, but she said,

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 probably not for all teachers.

According to the head mentor teacher, an effective principal who understands GE appears to be the most important distinction between schools that are doing very well with GE and those that are not. One characteristic of the most effective principals is that they find ways of removing ineffective or uncooperative teachers from their schools.

Outcomes for GE have not been rigorously evaluated. However, several teachers bragged during interviews that their own class-average test scores had risen by 30 or more percentiles—for example, from the 13th to the 45th, and so forth—in the space of 1 year. The key was GE’s apparently effective program of professional development. Just how effective the program has been over the long term is something that future evaluations may try to measure.

## 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 racial 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, and behaviors, or in any combination of the three.

Consider teacher perceptions of current student performance. If the benchmark for bias in teacher perceptions of current performance is unconditional racial neutrality, then 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 found no clear evidence of any type 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*. Conversely, 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 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."<sup>65</sup> 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 that teachers' beliefs can matter. Because potential is unobserved, racial bias of this type is virtually impossible to gauge with any reliability. Still, it seems especially likely that teachers underestimate the potential of students whose current performance is poor, including disproportionate numbers of Blacks. Similarly, 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 the top and the bottom of the test score distribution, stereotypes of Black intellectual inferiority are reinforced by past and present disparities in performance and probably cause teachers to underestimate Blacks' potential more than Whites'. If they expect that Black children have less potential, teachers probably search with less conviction than they should for ways of helping Black children to improve and 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 GE program with the assertion, "My kids couldn't do that." If they had gone on teaching the way they had always taught, that judgment would have been correct. But when they changed their teaching, they learned that they were wrong. Similarly, outside of GE, Guskey (1982) showed that teachers could learn responsive teaching methods that weaken the link between past and future performance. Teachers who are helped by professional development to improve their classroom practices can have "seeing is believing" experiences that challenge their prior biases. We need more research on how professional development programs affect both test score levels and the Black-White test score gap.

Even without the biases discussed above, teachers' beliefs probably affect Black students more than Whites. Evidence is quite thin, but the few studies that bear on this hypothesis appear to support it. Jussim et al. (1996) found that teachers' perceptions of sixth graders' math performance in October did not contain a racial bias once they controlled past performance and attitudes. Nevertheless, the impact of teachers' October perceptions on May math scores was almost 3 times larger for Blacks than for Whites. Further, the effect was 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. For example, Casteel (1997) found that Black eighth and ninth graders were more eager to please teachers, but Whites were more concerned about pleasing parents. 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 perhaps more deferential themselves to teachers. We do not know the answers. Future research should actively pursue these questions, including the implications for policy, teaching, and parenting.

My bottom-line conclusion requires speculation because the research is so incomplete. It 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. Unfortunately, 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 can establish that children of all racial and ethnic groups have more potential than most people have assumed. As the evidence accumulates, we should be able to focus with greater determination on cultivating and harvesting all that youthful minds embody.<sup>66</sup> 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.

### NOTES

1. See Phillips, Crouse, & Ralph (1998). The Black-White gap in skills at the beginning of primary school is smaller for more recent cohorts. Hedges & Nowell (1998) found evidence that reductions in Black-White test score disparity across cohorts during the 1970s and 1980s were due primarily to a narrowing of the Black-White gap in years of schooling among parents.

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 Phillips, Crouse and Ralph (1998) for a discussion of methodological decisions that determine whether the gap appears to be constant or widening with time within a cohort.

3. When possible, this article will present effect sizes measured in standard deviation units. For example, if one group of students experiences a particular treatment and an otherwise 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 example, for an outcome that is normally distributed, an effect size of .20 moves a student from the 50th to the 58th percentile; an effect size of .50 moves him or her to the 69th percentile; and an effect size of .80 moves the student to the 79th percentile.

4. Her comment is about racial disparity in ability-group assignments.

5. This immediate discussion concerns experiments. However, Type I benchmarks are also used sometimes in naturalistic settings. Specifically, unconditional racial neutrality may seem the only morally defensible alternative in the absence of reliable information *about individuals* on which to base a benchmark that is not unconditionally race neutral.

6. Baron, Tom, and Cooper (1985) reported that effect sizes could be retrieved for only 6 of the 16 studies. In these 6 studies, the Black-White differences in teacher expectations averaged half a standard deviation. If 9 of the other studies are assumed to have effect sizes of 0, and the 1 study with a significant result but no effect size is assumed to have an effect size of .36, then the average effect size across all 16 studies is .22.

7. Baron et al. (1985) conducted online computer searches of *Psychological Abstracts*, the Educational Research Information Center, and *Dissertation Abstracts International* as the basis for their initial bibliography. The descriptors that they used were crossings of the following terms: "teacher expect . . .," "teacher attitudes," or "teacher bias" with "racial," "ethnicity," "class," "socioeconomic," "social background," or "social characteristics."

8. To compute effect sizes for each outcome, the standard deviation among Blacks is used as the denominator because the pooled standard deviation was not given. The standard deviation among Whites is virtually the same as among Blacks. All effect sizes reported in this paragraph are calculated from numbers in DeMeis and Turner (1978), Table 2, p. 83. Effect sizes for the Black-White difference among students speaking standard English were 0.57, 0.52, 0.55, and 0.44 standard deviations respectively for "personality," "quality of response," "current academic abilities," and "future academic abilities." For Black English, the analogous numbers were 0.34, 0.44, 0.23, and 0.14. The fact that effect sizes were smaller for tapes on which students spoke Black English is not surprising, because speaking Black English would be an especially negative signal for a White student. Across the four outcomes, within-race effect sizes for Black English versus standard English ranged from 0.23 to 0.45 for Blacks and from 0.55 to 0.74 for Whites, always favoring standard English.

9. Whatever the reasons, average scores on standardized examinations tend to be lower for Blacks than for Whites. Hence, the most accurate prediction under most circumstances is that Whites will have higher scores. Moreover, given a group of Black and White students starting from equal baseline scores, the average future score among Whites in the group will tend to be closer to the overall future mean for Whites. The average among Blacks will tend to be closer to the overall mean for Blacks, which for most exams in most samples is lower than for Whites (Philips et. al., 1998).

10. See, for example, Eccles and Wigfield (1985) for a line of reasoning that might cause this to be true. Essentially, to deviate from a previously established trajectory, the student may need support from the teacher when the student begins the process of change. If the teacher continues treating the student as he or she did before the change, the student may decide that the environment is not sufficiently responsive to make the new goal attainable. The student feels a lack of control and returns to the old ways.

11. Experimental studies that expose teachers to different sequences of facts show that teachers' expectations are flexible to remain accurate as new information becomes available (e.g., see Shavelson, Cadwell, & Izu, 1977). The pattern of flexibility among teachers in real classrooms is not known.

12. For example, see the discussion herein of work by Guskey (1982), where improvements in teacher responsiveness *reduced* the accuracy of teachers' early predictions for end-of-semester performance.

13. This study covers 49 teachers and 934 fourth, fifth, and sixth graders in five cities across four census regions (see Haller, 1985, Footnote 4).

14. Discusses findings from Irvine (1985), which was a conference paper (see Irvine, 1990, p. 77).

15. The correlations between 2nd-week rankings and end-of-year test scores were .63 for White males and .62 for Black males. The correlation for Black males dipped by the 10th week but returned to the same range as for Whites by the end of the school year. Irvine (1990) emphasized the *difference* in the pattern for Black and White boys because the correlation dipped for Black boys in the 10th week. It seems to me, however, that the similarity is the more salient finding: Of three comparisons for boys and three for girls, in only one (boys at the 10th week) was the racial difference notable. Some teachers in Irvine's study were consistently more accurate than others. The teacher with the lowest correlations moved from .11 for the 2nd week to .56 for the end of the year. At the high end, one teacher had correlations of .91, .92, and .89 for the 2nd week through the end of the year.

16. Recall, for example, that two lines with different slopes each represent correlations of 1 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 in a

steeper line) even though the correlation between teachers' perceptions or expectations and end-of-year performance was the same for both groups. This possibility of different slopes is explored in Jussim, Eccles, and Madon (1996) and discussed herein, after discussing the other two types of racial bias.

17. Members of the task force that produced the report included a number of noted Black scholars: Sara Lawrence Lightfoot of Harvard, who was the principal author; James P. Comer of Yale; John Hope Franklin of Duke; and William Julius Wilson, then of the University of Chicago.

18. Of course, goal setting is affected by more than teachers' expectations of student potential. The curricular materials that are handed down to teachers from the administration and students' actual behaviors matter as well.

19. See Ford (1996) for useful discussions of this issue referring to theories of multiple intelligences. Also consider, using the terminology of the present article, that people who think that potential is very distinct from performance and that ability is equally distributed among the races will favor a Type I proxy for Type III. Others might favor a Type II proxy, augmented perhaps by a positive shift factor for all students, if they think that racial differences in performance are good approximations of racial differences in potential. Of course, it is difficult to find anyone willing to suggest this in print.

20. Miller (1995) presents a useful review of trends in surveys regarding beliefs about Black intellectual inferiority. He pointed 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.

21. Across different surveys, it is clear that fewer people answer yes to questions about intellectual inferiority of Blacks when the question asks explicitly about "in-born" or genetic differences (see the discussion in Miller, 1995, chapter 8).

22. The phrase "rumors of inferiority" comes from the title to Howard and Hammond (1985).

23. Leading researchers on teacher expectations claim that sustaining expectations are probably more prevalent than self-fulfilling expectations (e.g., see Good, 1987).

24. The most frequent explanation for failure is that teachers do not believe the information about the students. Sometimes teachers figure out the purpose of the experiment. Other times, teachers have their own sources of credible information or they have known the students long enough to form opinions before the experiment begins. In a meta-analysis of 18 experiments in which IQ or a similar measure of ability was the outcome, Raudenbush (1984) showed very clearly that studies finding evidence of the effect were primarily those where teachers had no opportunity to form an independent impression of students before the experiment began.

25. Smith (1980) did not say what percentage of the effect sizes she reported use the standard deviation of test score levels as opposed to test score gains in reporting effect sizes.

26. Actually, Brophy (1985) said that teachers' expectations probably make about a 5% difference, but he did not say whether he was talking about the difference in gain during a school year or the difference in total achievement (as in the level on a test). Brophy's statement was based on his own review of the literature where the period of individual studies is seldom more than a single school year.

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

28. It is not unusual, for example, for a teacher to say, "Betty is doing well now because she is repeating the grade and has seen some of this material before. I don't expect that she will do as well for the rest of the year." This teacher might be accurate in his or her current evaluation of the student, but still biased in his or her expectation. Other examples could reverse the expectation and make it more positive than current performance. There too, the expectations might or might not be biased when judged from the perspective of what past performance and attitudes would predict.

29. Also see Jussim (1989) and Jussim and Eccles (1992) for previous studies using the same data to study the accuracy of teachers' expectations, but not emphasizing racial differences. They speculated that the void in the literature stems from the political risk of studying groups who actually do differ. Researchers have avoided the risk by focusing on experimental studies that assume away differences.

30. Of the 1,664 sixth graders, 76 were African Americans. Ideally, one would want a larger number of African Americans in such a study. It is not clear why this data set contained so few. The data came from the Michigan Study of Adolescent Life Transitions, which was not initially designed to study racial differences. For more details about the Michigan Study of Adolescent Life Transitions, see Wigfield, Eccles, Maclver, Reuman, and Midgley (1991).

31. The study does not report raw means by race, but it does report that the correlations of race with grades and standard test scores were  $-.12$  and  $-.14$ , respectively (both with  $p < .001$ ), with Blacks having the lower scores and grades. These are probably smaller than in the typical national sample of Black and White sixth graders.

32. Jussim, Eccles, & Madon (1996) did separate calculations to measure whether the residual variance in teachers' perceptions left unexplained by the background factors was similar for Blacks and Whites. It was slightly higher for Blacks, but only 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  $.06$  ( $p < .05$ ),  $.07$  ( $p < .05$ ), and  $-.02$  (not significant) for performance, talent, and effort, respectively" (p. 355). Although 2 of the 3 are statistically significant, these suggest only a very small difference in accuracy, with less accuracy for Blacks than for Whites.

Regarding social class, there was a small positive relationship found between parents' education and the teacher's perception of the student's talent. There were also some small gender effects. Teachers' perceptions of performance and effort were higher for girls after controlling for the factors listed in the text. Hence, it appears that teachers were relying to a small extent on a gender stereotype, though not necessarily a false one.

33. As background variables, the equation to predict scores and grades includes race, math grades from the end of fifth grade, math scores from the end of fifth or beginning of sixth grade, self-concept of ability, self-reported effort at math, self-reported time spent on homework, and indices for the intrinsic and extrinsic value of math to the student. Interactions of student's race with teacher perceptions of effort and talent were tried, but these produced strange results because of collinearity with the interaction for race and performance. The result for performance might best be understood to represent the interaction of race with all aspects of the teacher's perceptions.

34. For Michigan Educational Assessment Program scores, the effect size was only  $.14$  for Whites, but it was  $.37$  for African Americans ( $p < .001$ ). This effect size for Whites is quite close to the effect size of  $.17$  for math achievement scores that Smith (1980) reported in her meta-analysis. For grades, the effect size for African Americans was  $.56$  compared with  $.20$  for Whites ( $p < .01$ ). Calculations conducted by Jussim after the article was published use a specification that includes additional interaction terms. That specification adds interactions

of race with the student's past grades and scores. The effect size for Michigan Educational Assessment Program scores rose from the original .37 to become .58 for African Americans and dropped from .14 to .13 for Whites. With the additional interaction terms, the coefficients on past grades and scores were estimated to be somewhat smaller for African Americans than for Whites. Hence, it appears that Blacks' performance was more dependent than Whites' on current teachers' opinions and less anchored in measures of past performance. One might speculate that this is because past grades and scores for Black students were less accurate than for Whites as measures of their knowledge or potential, but we cannot know from the information available.

35. It is not clear what to make of the lack of relationship between self-reports of effort and teachers' perceptions of effort. If teachers are really grossly inaccurate in their assessments of effort, this could contribute to the disengagement of children who are believed not to be trying when they really are. It could also contribute to a lack of challenge for students who are slacking off when they appear to be working hard and actually would be well served by being asked to do more.

36. The fact that teacher perceptions were also stronger predictors for females and for Whites from low-income households makes it more likely that this is a real effect for Blacks. Further, because Jussim et al. (1996) found no unexplained racial differences in 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 mean very different things or have different implicit scalings for Blacks and Whites. Still, these results need to be replicated several times in future research to be firmly established.

37. Jussim et al. (1996) assumed that current perceptions were good estimates of expectations for the future, and they used the words *perceptions* and *expectations* interchangeably. Jussim (personal communication, 1998) has argued that this is a reasonable assumption based on other research regarding the processes by which people form expectations. This, however, might be inappropriate for the reasons explained in the text. I have not found any research with data on teachers' perceptions of current performance along with expectations for future performance for both Blacks and Whites.

38. The sample included 928 Whites and 761 African Americans from nine schools and 12 classes in two public school districts.

39. See the discussion in pages 46-49. Irvine cites two dissertation studies from the 1970s and one journal article from the late 1960s when she writes, "Researchers have found that black and other minority pupils are more negatively affected by teacher expectations than white students are" (p. 49). The studies she cites to make this point are Baker (1973), Krupczak (1972), and Yee (1968). I have not found any studies that contradict these few.

40. Jussim, Eccles, & Maddon (1992) cited an earlier paper by Steele (1996).

41. 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 Black. Although I did not believe that the pattern was unfair, I wanted the teacher to know and I wanted to know for myself that I was an exception to it. 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 establish me as an exception. 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 the racial pattern in group memberships remained.

42. These fourth- and sixth-grade results are from a single year for two different cohorts; hence, although the differences appear consistent with a trend, they do not clearly establish it. Also, Gross (1993) did not give sample sizes broken down by the number who were above, at, or below grade levels. Hence, we are not sure how many children each of these numbers represents.

43. Gross (personal communication, February 1998) reported that she conducted regression analyses using students' classroom performance as predictors, and the CATM test performances of the Black high achievers were below what was predicted. However, these calculations were unpublished and, because they are a decade old, are no longer available.

44. Mickelson (1990) reported (but did not show) that results with standardized test scores as the dependent variables produced the same story as for grades. In predicting grade point average, standardized coefficients on concrete attitudes in the full specification were .111 ( $p < .05$ ) for Blacks and .190 ( $p < .01$ ) for Whites. These are not large effects. Still, the fact 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 in Table 2, provide important insight. The coefficient for abstract attitudes was about a fifth as large as for concrete attitudes. The t-statistic on the coefficient for abstract attitudes was about 1 for Whites and less than .5 for Blacks. The regressions control for: mother's and father's occupations and educations, a locus of control index, the student's weekly hours worked, the percentage of close friends planning to attend a 4-year college, and an indicator variable for each of the eight schools in the sample.

45. Mickelson (1990) studied 1,193 seniors from eight high schools in the Los Angeles area during the 1983 school year. She analyzed only the responses from Blacks and Whites, who comprised 41% and 59%, respectively, of her working sample. Mickelson's indices measured levels of agreement with the following statements. For each index (abstract and concrete), higher levels of agreement correspond to higher values.

#### Abstract Attitudes

1. Education is the key to success in the future.
2. If everyone in America gets a good education, we can end poverty.
3. Achievement and effort in school lead to job success later on.
4. The way for poor people to become middle class is for them to get a good education.
5. School success is not necessarily a clear path to a better life.
6. Getting a good education is a practical road to success for a young Black [White] man [woman] like me.
7. Young White [Black] women [men] like me have a chance of making it if we do well in school.
8. Education really pays off in the future for young Black [White] men [women] like me.

#### Concrete Attitudes

1. Based on their experiences, my parents say people like us are not always paid or promoted according to our education.
2. All I need to learn for my future is to read, write, and make change.
3. Although my parents tell me to get a good education to get a good job, they face barriers to job success.
4. When our teachers give us homework, my friends never think of doing it.
5. People in my family have not been treated fairly at work, no matter how much education they have.
6. Studying in school rarely pays off later with good jobs.

I think that to call these attitudes “concrete” is a gross misnomer and that it has confused the interpretation of Mickelson’s (1990) work. They are just as “abstract” as the others. What they measure is ambivalence or doubt about the first list, but “fairness” is not “concrete.”

46. The “middle class” or “working class” designation is based on a combination of standard blue/white collar distinctions and the parents’ level of education.

47. Also see Babad (1980, 1985) for related research from Israel.

48. The possibility that confederate students’ behaviors confounded findings of racial bias in Rubovits and Maehr’s (1973) study led Taylor (1979) to try a different experimental design. She placed “phantom” 6-year-old students behind a one-way glass where they could allegedly see the teacher and respond to the teacher’s questions and instructions. This removed any effects of targets’ actual behaviors. The teachers were told that the students could see and hear them and would respond to their instructions by pushing buttons that lighted up on a panel that the teachers could see.

Teachers in Taylor’s (1979) study were 105 White female undergraduates at the University of Massachusetts, Amherst. They were studying education, just as in Rubovits and Maehr (1973). 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 showed that the teachers had believed this premise. Each was told in written instructions that the student was Black or White, male or female, and of high or low ability. In fact, unknown to the participants, all of the “student feedback” during the experiment was provided by a single adult who was hidden from view and gave answers by manipulating 10 lights on a panel. The person giving the answers was blind to what any given teacher had been told about the “student.” Sessions were videotaped and teachers’ behaviors were coded.

49. There are many sources from which to draw this standard finding. For example, Irvine (1990), in Table 3.3, tabulated studies conducted in naturalistic classroom settings. Of 17 findings from 16 studies, Whites were favored in 9 cases, there was no difference in 4 cases, the opposite race of the teacher was favored in 2 cases, the same race as the teacher was favored in 1 case, and Blacks were favored in 1 case. All but 1 of the studies on the list are from the 1970s and early 1980s; 1 is from 1969. For a discussion of earlier studies, see Brophy and Good (1974).

50. Blacks rate lower by a statistically significant margin for each of the three variables in the index.

51. See Brophy and Good (1974) and other literature that they discuss regarding the types of students that teachers like to teach. Also see below in the present article.

52. Teachers’ expectations might be less flexible for Black students than for Whites (though we do not know for sure that they are). Inflexible teacher perceptions for Blacks might lead to teacher behaviors that reinforce problem behaviors for low-performing students and promote good behaviors for high-performing students. This is a ripe topic for future research.

53. See the Ferguson (1998) for a discussion of how teacher’s race and social class background might affect student performance. Simply matching the race of the student and the child is too simple a prescription because social class and professional competence also appear to be important.

54. One important thing to communicate about academic ability is that it is not immutable—that sustained effort mobilizes and develops ability. Dweck (1991) and Dweck and Leggett (1988) found statistically significant evidence among White children that those believing ability is fixed tend to adopt performance goals, whereas those believing otherwise tend to adopt mastery goals. Among those who believe ability is fixed and that their own

endowment is low, the performance goal is to hide their ignorance. Among those who believe ability can be mobilized and developed by sustained effort, both high and low achievers tend toward mastery goals. The same is probably true for Blacks, but the research that would show it has not, to my knowledge, been done.

55. See the literature review section of Cook and Ludwig (1998).

56. In a summary of the literature on wait time, Rowe (1986) reported,

[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* [italics added]. They did more task relevant talking and took a more active part in discussions than they had before. (p. 45)

There are other references in Rowe's (1986) paper to the studies that develop these findings. Wait time is shorter for low-performing students.

57. Studies show mixed results regarding techniques for improving corrective feedback. Some are more successful than others. See, for example, Slavin's (1987) review of the literature of mastery learning.

58. Early in the semester, teachers classified each student into one of five groups of equal size based on "probable achievement" in the course. At the end of the semester, they repeated this exercise, placing students in five groups of equal size based on "achievement potential."

59. Brattesani, Weinstein, and Marshall (1984) found strong evidence that standardized test scores were predicted less well by past performance in classrooms where there was more differential treatment. The point they made in the paper is that teacher expectations become self-fulfilling prophecies only when communicated through differential treatment. They found that teacher expectations are stronger predictors in classrooms with more differential treatment. However, because the teachers' expectations were collected in April of the school year, I regard them as reports from the end of the school year, not self-fulfilling predictions.

60. See Table 4 in Brattesani et al. (1984). The sample sizes are small, so the difference in percentages of low achievers making large gains does not reach statistical significance. Nevertheless, the magnitudes of the gains are large among the small number of students involved. For interesting related work, see Weinstein, Marshall, Sharp, and Botkin (1987), who showed that even first graders can accurately report differential treatment by teachers of their peers, but it is not until third and fifth grade that their reports regarding their own differential treatment become more accurate.

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

62. Most of the time, she maintained, the resistance could be reduced by a combination of two responses. First, she would assure the teacher that she could slip into GE gradually, implementing some of it first and other things later. Second, she would model the GE practices to which the teacher was resistant. She would do so at that teacher's school, preferably in that teacher's classroom, and always with the greatest respect and tact. When a teacher witnessed a mentor teacher demonstrating a method with the teacher's own students, and succeeding, he or she usually became (or claimed to become) more open to giving it a try.

63. In one worst case example, a teacher had failed with a new method during the year before the introduction of GE. Several other teachers at her school were using the GE method and doing well. She, however, was sure that GE could not work for her. So, she continued teaching one letter per week to her first-grade class of "slow learners," with no pressure from her passive principal to change.

64. The names of the teachers and schools have been changed, but the facts and quotations are real.

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

66. Of course, in addition to strong leadership and professional development, stronger performance incentives for teachers, no matter what their expectations, should be part of the expression of society's determination. The search for ways of designing and implementing such incentives is currently quite active. Three recent books with chapters that emphasize incentives and accountability are Hanushek and Jorgenson (1996), Hanushek (1994), and Ladd (1996).

## REFERENCES

- Babad, E. Y. (1980). Expectancy bias in scoring as a function of ability and ethnic labels. *Psychological Reports, 46*, 625-626.
- Babad, E. Y. (1985). Some correlates of teachers' expectancy bias. *American Educational Research Journal, 22*(Summer), 175-183.
- Baron, R., Tom, D. Y. H., & Cooper, H. M. (1985). Social class, race and teacher expectations. In J. B. Dusek (Ed.), *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum.
- Brattesani, K. A., Weinstein, R. S., & Marshall, H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology, 76*, 236-247.
- Brophy, J. (1985). Teacher-Student interaction. In J. B. Dusek (Ed.), *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum.
- Brophy, J., & Good, T. (1974). *Teacher-Student relationships: Causes and consequences*. New York: Holt, Rinehart & Winston.
- Cabello, B., & Burstein, N. D. (1995). Examining teachers' beliefs about teaching in culturally diverse classrooms. *Journal of Teacher Education, 46*(September/October), 285-294.
- Casteel, C. (1997). Attitudes of African American and Caucasian eighth grade students about praises, rewards, and punishments. *Elementary School Guidance and Counseling, 31*(April), 262-272.
- Coates, B. (1972). White adult behavior toward Black and White children. *Child Development, 43*, 143-154.
- Cook, P. J. & Ludwig, J. (1998). The burden of 'acting White': Do Black adolescents disparage academic achievement? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap*. Washington DC: Brookings Institution Press.
- DeMeis, D. K., & Turner, R. R. (1978). Effects of students' race, physical attractiveness, and dialect on teachers' evaluations. *Contemporary Educational Psychology, 3*, 77-86.
- Dweck, C. (1991). Self-Theories and goals: Their role in motivation, personality and development. In R. A. Dienstbier (Ed.), *Nebraska Symposium on Motivation, 1990*. Lincoln, NE: University of Nebraska Press.
- Dweck, C., & Leggett, E. L. (1988). A social cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.
- Eccles, J., & Wigfield, A. (1985). Teacher expectations and student motivation. In J. B. Dusek (Ed.), *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum.
- Egan, O., & Archer, P. (1985). The accuracy of teachers' ratings of ability: A regression model. *American Educational Research Journal, 22*, 25-34.

- Entwisle, D. R., & Alexander, K. L. (1988). Factors affecting achievement test scores and marks of Black and White first graders. *The Elementary School Journal*, 88(5), 449-471.
- Evertson, C., Brophy, J., & Good, T. (1972). *Communication of teacher expectations: First grade* (Report No. 91). Austin: University of Texas at Austin, Research and Development Center for Teacher Education.
- Feldman, R. S., & Orchowsky, S. (1979). Race and performance of student as determinants of teacher nonverbal behavior. *Contemporary Educational Psychology*, 4, 324-333.
- Ferguson, R. F. (1993). *Spreading the paradigm of a master teacher: The Great Expectations initiative in Oklahoma* (Working Paper of the Taubman Center for State and Local Government). Cambridge, MA: Harvard University, John F. Kennedy School of Government.
- Ford, D. Y. (1996). *Reversing the underachievement among gifted Black students*. New York: Teachers College Press.
- Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the "burden of 'acting White.'" *The Urban Review*, 18(3), 176-206.
- Gaines, M. L. (1990). *Accuracy of teacher prediction of elementary student achievement*. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED320942)
- Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education*, (4), 32-47.
- Gross, S. (1993). Early mathematics performance and achievement: Results of a study within a large suburban school system. *Journal of Negro Education*, 62, 269-287.
- Guskey, T. R. (1982). The effects of change in instructional effectiveness on the relationship of teacher expectations and student achievement. *Journal of Educational Research*, 75, 345-348.
- Haller, E. J. (1985). Pupil race and elementary school ability grouping: Are teachers biased against Black children? *American Educational Research Journal*, 22(4), 465-483.
- Hanushek, E. A. (1994). *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution.
- Hanushek, E. A., & Jorgenson, D. W. (Eds.). (1996). *Improving America's schools: The role of incentives*. Washington, DC: National Academy Press.
- Herrnstein, R. J. & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hoge, R., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement level. *Journal of Educational Psychology*, 76, 777-781.
- Howard, J., & Hammond, R. (1985, September). Rumors of inferiority. *New Republic*, 72, 9.
- Irvine, J. J. (1985). *The accuracy and stability of teachers' achievement expectations as related to students' race and sex*. Paper presented at the meeting of the American Educational Research Association, Chicago.
- Irvine, J. J. (1990). *Black students and school failure: Policies, practices, and prescriptions*. Westport, CT: Greenwood.
- Jencks, C. & Phillips, M. (Eds.). 1998. *The Black-White test score gap*. Washington DC: Brookings Institution Press.
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, 57, 469-480.
- Jussim, L., & Eccles, J. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63, 947-961.
- Jussim, L., Eccles, J., & Madon, S. (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.

- Kleinfeld, J. (1972). The relative importance of teachers and parents in the formation of Negro and White students' academic self-concepts. *Journal of Educational Research*, 65, 211-212.
- Kluegel, J. R. (1990). Trends in Whites' explanations of the Black-White gap in socioeconomic status, 1977-1989. *American Sociological Review*, 55(August), 512-525.
- Ladd, H. F. (Ed.). (1996). *Holding schools accountable: Performance-Based reform in education*. Washington, DC: Brookings Institution.
- Lightfoot, S. L. (1978). *Worlds apart: Relationships between families and schools*. New York: Basic Books.
- Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review*, 8, 193-210.
- Mickelson, R. A. (1990). The attitude-achievement paradox among Black adolescents. *Sociology of Education*, 63, 44-61.
- Miller, L. S. (1995). *An American imperative: Accelerating minority educational advancement*. New Haven, CT: Yale University Press.
- Mitman, A. (1985). Teachers' differential behavior toward higher and lower achieving students and its relation to selected teacher characteristics. *Journal of Educational Psychology*, 77, 149-161.
- Monk, M. (1983). Teacher expectations? Pupil responses to teacher mediated classroom climate. *British Educational Research Journal*, 9, 153-166.
- Ogbu, J. (1978). *Minority education and caste: The American system in cross-cultural comparison*. New York: Academic Press.
- Ogbu, J. (1983). Minority status and schooling in plural societies. *Comparative Education Review*, 27, 168-203.
- Ogbu, J. (1987). Opportunity structure, cultural boundaries, and literacy. In J. Langer (Ed.), *Language, literacy, and culture: Issues of society and schooling*. Norwood, NJ: Ablex.
- Pedulla, J., Airasian, P., & Madaus, G. (1980). Do teacher ratings and standardized test results of students yield the same information? *American Educational Research Journal*, 17, 303-307.
- Phillips, Crouse, & Ralph. (1998). In C. Jencks (Ed.), *The Black-White test score gap*. Washington, DC: Brookings Institution.
- Puma, M., Jones, C., Rock, D., & Fernandez, R. (1993). *Prospects: The congressionally mandated study of educational growth and opportunity* (Interim Report prepared for the U.S. Department of Education, Planning and Evaluation Service).
- Raudenbush, S. 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, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3(6).
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37(1), 43-50.
- Rubovits, P. C., & Maehr, M. L. (1973). Pygmalion Black and White. *Journal of Personality and Social Psychology*, 25, 210-218.
- Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. *American Educational Research Journal*, 14(Spring), 83-97.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175-213.

- Smith, M. L. (1980). Teachers' expectations. *Evaluation in Education*, 4, 53-56.
- Solorzano, D. (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.
- Steele, C. M. (1992, April). Race and the schooling of Black Americans. *Atlantic Monthly*, 68-72.
- Steele, C. M. & Aronson, J. (1998) Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap*. Washington DC: Brookings Institution Press.
- Taylor, M. C. (1979). Race, sex, and the expression of self-fulfilling prophecies in a laboratory teaching situation. *Personality and Social Psychology*, 6, 897-912.
- Weinstein, R. S. (1985). Student mediation of classroom expectancy effects. In J. B. Dusek (Ed.), *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum.
- Weinstein, R. S., Marshall, H. H., Sharp, L., & Botkin, M. (1987). Pygmalion and the student: Age and classroom differences in children's awareness of teacher expectations. *Child Development*, 58, 1079-1092.
- Wigfield, A., Eccles, J. S., MacIver, D., Reuman, D., & Midgley, C. (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, 552-565.
- Willis, S. (1972). *Formation of teachers' expectations of students' academic performance*. Unpublished doctoral dissertation, University of Texas, Austin.
- Willis, S., & Brophy, J. (1974). The origins of teacher attitudes toward young children. *Journal of Educational Psychology*, 66(4), 520-529.

*Ronald F. Ferguson is an economist and senior research associate at the Wiener Center for Social Policy. He has taught at Harvard University's John F. Kennedy School of Government since 1983. He participates in a variety of consulting and policy advisory activities on issues of education, employment, youth development, and urban development.*