SOCIAL PERCEPTION, SOCIAL STEREOTYPES, AND TEACHER EXPECTATIONS: ACCURACY AND THE QUEST FOR THE POWERFUL SELF-FULFILLING PROPHECY

Lee Jussim
Jacquelynne Eccles
Stephanie Madon

I. Introduction

How are social perception and social reality related? Social psychology has long emphasized the power of beliefs to create reality, and the power of interpersonal expectancies to create social problems (e.g., Gage & Cronbach, 1955; Jones, 1986, 1990; Merton, 1948; Miller & Turnbull, 1986; Snyder, 1984). Social scientists have a longstanding interest in one particular source of expectations—stereotypes—largely because stereotypes may contribute to social inequalities and injustices. But are people so malleable that they readily fulfill others' inaccurate expectations? How accurate are interpersonal expectations? To what extent do stereotypes bias person perception and lead to self-fulfilling prophecies? Who is most vulnerable to self-fulfilling prophecies?

In this article, we address these questions as follows. First, we present a brief overview of research on accuracy, error, bias, and self-fulfilling prophecies. Second, we review our own research showing that teacher expectations predict student achievement mainly because they are accurate although they do lead to small self-fulfilling prophecies and biases. We subsequently embark on a quest to identify conditions under which self-fulfilling prophecies might be considerably more powerful. Third, therefore, we report the results of new research showing that teacher expectancies effects are more powerful among girls, students from lower socioeconomic status (SES) backgrounds, or African-Americans.
Social psychological research on stereotypes suggested a possible explanation for this pattern: Teachers rely on stereotypes in developing expectations for students from stigmatized groups, and because such expectations will often be inaccurate, they are also more likely to be self-fulfilling. Therefore, we review the general literature on the role of stereotypes in creating self-fulfilling prophecies and on issues of accuracy and inaccuracy in stereotypes more generally.

We then address some of these issues empirically in two studies that examined whether teacher perceptions of differences among students belonging to different demographic groups (boys or girls, middle class or poor, African-American or White) were biased or accurate. Although these studies provided some evidence of bias (surprisingly, these biases were usually in favor of students from culturally stigmatized groups), they also showed that, in general, differences in teachers' perceptions of students from the different groups corresponded well to actual differences between those same groups of students.

Although we found such results particularly interesting in light of the social sciences' emphasis on inaccuracy of stereotypes, they left us still unable to explain why self-fulfilling prophecies were stronger among students from stigmatized groups. We therefore speculate that students who feel devalued in school will be particularly susceptible to confirming teachers' expectations. Although we cannot test this idea directly, we summarize another new study that provides indirectly supportive evidence by showing that students with low self-concepts of ability or with previous records of low achievement were, much like students from stigmatized groups, considerably more vulnerable to self-fulfilling prophecies. We conclude this chapter by reviewing other moderators of expectancy effects, discussing the evidence showing whether self-fulfilling prophecies accumulate or dissipate over time, and making recommendations for future research on self-fulfilling prophecies.

II. Accuracy, Error, Bias, and Self-Fulfilling Prophecy

A. A BRIEF OVERVIEW

Research on accuracy, error, bias, and self-fulfilling prophecy have long traditions in social psychology. Error and bias research dates back at least to the emphasis in the 1930s on the inaccuracy of stereotypes (e.g., Katz & Braly, 1933; LaPiere, 1936). The idea that many social injustices and inequalities reflect self-fulfilling prophecies was first suggested in the 1940s (Merton, 1948). There was lively interest in accuracy through the 1950s (e.g., Taft, 1955; Vernon, 1933), which came to an abrupt and premature end after Cronbach identified many seemingly difficult statistical and methodological problems involved in assessing certain kinds of accuracy (Cronbach, 1955; Usser & Cronbach, 1955). At the same time, the New Look in Perception, which emphasized a myriad of ways in which perceivers' goals, needs, fears, and motives could influence and undermine the veridicality of perception, initiated a revolution in approaches to perception, at least in social and personality psychology (e.g., Allport, 1955; Bruner, 1957).

Subsequently, from the 1960s through much of the 1980s, social and personality psychology emphasized a host of errors and biases in social perception (e.g., Kahneman & Tversky, 1973; Miller & Tversky, 1986; Nisbett & Ross, 1980). Central to this effort was the work by Rosenthal demonstrating that experimenters and teachers can evoke expectancy-confirming behavior from both animals and people (see Rosenthal, 1974, for a review; Rosenthal & Jacobson, 1968; see Rosenthal & Rubin, 1978, for a meta-analysis). Numerous researchers then followed up this work with studies of the potentially self-fulfilling effects from expectancies of all sorts in and out of the laboratory (see reviews by Jones, 1977; Junlin, 1986; Snyder, 1984).

Although a few researchers did attempt to keep the study of accuracy alive after the 1960s (Archer & Akert, 1977; McCauley & Stitt, 1978), they were rare voices barely heard above the din of the zeitgeist emphasizing error, bias, and self-fulfilling prophecy. In the 1980s, however, four articles sparked the beginning of a renaissance in interest in accuracy. McArthur and Baron (1983) presented the first coherent theoretical alternative to the constructivist zeitgeist that had dominated thinking about social perception for 30 odd years. They took the ecological approach, which was originally developed to study object perception (Gibson, 1979), and applied it to social perception. This theory emphasized the information in the stimulus, which was in sharp contrast to the social cognitive emphasis on the categories, prototypes, schemas, and assorted cognitive structures existing in the perceivers' mind.

Next, Swann (1984) presented a broad and sweeping review of research on accuracy. Perhaps most influential was his discussion of "circumscribed" accuracy. Swann (1984) argued that perceivers often have no interest in predicting the behavior of targets across all situations and for all time. Thus, it is inappropriate to hold them to this standard. Instead, he suggested that people are usually content to understand and predict others' behavior only when they interact with those others. In terms of this more circum-
scribed notion of accuracy, Swann (1984) speculated that people might be considerably more accurate than had previously been recognized.

Next, Funder (1987) presented a conceptual and empirical assault on what he believed was social psychology's misplaced emphasis on error and bias. He focused on two main points: 1) Social psychology's knowledge base regarding error and bias stemmed almost exclusively from studies that were originally designed to assess social perceptual processes, and 2) the studies assessing error did not address the accuracy of outcomes produced by such processes. Funder thus drew a parallel between the laboratory social cognitive work on illusions, biases, and errors, and the laboratory social perception research on illusions. Researchers used controlled visual illusions to probe the dynamics of social information processing; they never assumed that these illusions reflected deficiencies likely to occur in perception under normal conditions. This, Funder argued, was also the most appropriate interpretation for research on human judgment. For Funder, accuracy was an issue of content, not of process. He also presented data documenting people's moderate to strong accuracy in perceiving others' personalities.

At about the same time, Kenny was publishing numerous articles describing the social relations model (Kenny, 1994 for a review). In one of the most influential of these early articles, Kenny and Albright (1987) explained how the social relations model could be used to generate error and accuracy in social perception; 2) pointed out the similarity between the accuracy component assessed by the social relations model and Cronbach's components of accuracy; and 3) showed how, when applied to social interaction, the model empirically documented considerable accuracy in social perception.

By 1990, the accuracy dijinn was most of the way out of the box. One more paper popped the cork completely. A main bastion of the scholarly emphasis on error and bias was the expectancy effects literature, especially the literature on social stereotypes (see, e.g., the strong emphasis on the power of expectations to create reality in reviews by Jones (1990), Miller & Turnbull (1986), and Snyder (1984)). In contrast, Jasmin (1991) argued that, on the evidence available, there was no empirical basis for this emphasis. He presented a model showing how people's beliefs could be in touch with reality more of the time, and yet still sometimes produce biases in person perception leading to self-fulfilling prophecies. This model was then used to interpret previous research on the effects of interpersonal expectations. Jasmin concluded that 1) interpersonal expectations can lead to biases and self-fulfilling prophecies, but these effects tend to be quite small; 2) perceivers' expectations of targets' future behavior and their impressions of targets' past behavior tend to be reasonably accurate; and 3) the evidence on the accuracy of social stereotypes is quite mixed (some accuracy, some inaccuracy). Jasmin also showed that much of what looked like expectancy-induced bias in experimental laboratory studies could actually enhance people's perceptual accuracy under some naturally occurring conditions.

The revival in interest in accuracy has, however, with a few exceptions, occurred in parallel with continued interest in error and bias. Many researchers still study and emphasize error, bias, or self-fulfilling prophecy (e.g., Gilbert & Malone, 1965; Snyder, 1972; Stangor, 1995); others focus primarily on accuracy (e.g., Ambady & Rosenthal, 1992; Bourke & Lichter, 1992; Funder & Cohen, 1998; Levenson & Kenny, 1993). Although some researchers have attempted to integrate accuracy and bias (e.g., Brewer, 1988; Fiske & Neuberg, 1990; Higgins & Berscheid, 1987; Kunda, 1990), most have relied primarily on experimental laboratory studies (see Kenny, 1994, for a review of partial exceptions—nonexperimental laboratory studies of accuracy and bias). These attempts, however, suffer from the conceptual problem first identified by Funder (1987): Because they focus on process instead of content, and because their relevance to naturalistic situations is unclear, they provide little empirical evidence on accuracy, error, bias, and self-fulfilling prophecy in daily life. (Even some hardcore experimental social psychologists have expressed sympathy with the view that the relevance of laboratory studies to daily life is an unanswered, open, empirical question.) (See Gilbert & Malone, 1995, p. 35.)
B. THREE SOURCES OF EXPECTANCY CONFIRMATION

Perceivers' expectations may be confirmed for any of at least three reasons: two involve expressions on behavior or perceptions and one that does not. First, perceivers' expectations sometimes produce self-fulfilling prophecies: Their initially erroneous expectations may cause targets to act in ways consistent with those expectations. (Casterline, 1979; Darley & Pizzolo, 1980; Jussim, 1986; Rosenthal & Jacobson, 1968). Second, expectations may lead to perceptual biases: perceivers may interpret, remember, and/or explain targets' behavior in ways consistent with their expectations. This type of expectancy confirmation exists in the minds of the perceiver rather than in the behavior of the target (Darley & Pizzolo, 1980; Eccles & Jacobs, 1986; Jussim, 1991; Miller & Turnbull, 1988). Self-fulfilling prophecies and perceptual biases both represent perceivers' expectations creating (or "constructing") social reality, either creating an objective social reality (when self-fulfilling prophecies change targets' actual behavior) or a subjective social reality (when perceptual biases influence perceivers' evaluations of target behavior). Finally, in contrast, expectations also may accurately reflect or predict social reality without influencing either objective target behavior or even subjective perceptions of that behavior (Gregory, 1985; Jussim, 1991).

Although these three expectancy phenomena are conceptually distinct, they are not mutually exclusive. Any combination of the three (or none at all) can characterize relations between perceivers' expectations and student achievement (Jussim, 1989, 1991; Jussim & Eccles, 1992). Consider a teacher who believes a student is especially bright. The teacher may be accurately perceiving—this student may indeed have a stronger academic background than most others. Furthermore, highly positive interactions with the teacher may lead the student to achieve even more highly—this, demonstrating a self-fulfilling prophecy. Finally, perceptual biases may lead the teacher to evaluate the student even more favorably than is warranted by the student's objective performance.

Although expectations may lead to many combinations of self-fulfilling prophecy, perceptual bias, and accuracy, they may also lead to none; expectations can be both inaccurate and nonfunctional. For example, a teacher may expect a student to be a low achiever. Nevertheless, this student may successfully complete his/her homework assignments in a timely and thorough manner and go on to perform above average on a highly credible standardized achievement test and receive modest "As" on in-class tests. The teacher may simply acknowledge the error (i.e., the original expectation was erroneous, but there is no perceptual bias), and the student may continue to perform highly (no self-fulfilling prophecy).

Although this article focuses exclusively on relations between teacher expectations and student achievement, expectancy effects undoubtedly occur in many other relationships employer-employee, therapist-client, and parent-child. Consequently, as we analyze ways to distinguish among self-fulfilling prophecies, perceptual biases, and accuracy, and examine processes underlying expectancy-related phenomena, our discoveries may have some relevance and applicability to many other relationships beyond teachers and students. (see also Eccles et al., 1995; Eccles & Hoffman, 1984; Jacobs & Eccles, 1992; Jussim, 1990, 1991; Jussim & Eccles, 1995; Jussim & Fleming, in press).

III. Teacher Expectations

There are few contexts more important for investigating self-fulfilling prophecies than teachers' expectations for their students. Ever since Rosenthal and Jacobson's (1968) seminal and controversial (e.g., Elashoff & Snow,
7917) Pygmalion study, writers to such scholarly journals and the popular press have implicated teacher expectations as a major perpetrator of injustices and inequalities based on ethnicity,2 social class, and gender (see Winsberg, 1987, for a review). In this article, we present evidence suggesting that such claims present an oversimplified and exaggerated picture of the role of teacher expectations in perpetuating social inequalities. This evidence will convey two main points. First, we briefly review our own and others' research documenting that naturally occurring teacher expectations generally lead to only small self-fulfilling prophecies and perceptual biases. This research also shows that teacher expectations predict student achievement primarily because they are accurate. Even though teacher expectation effects are generally small, under some conditions or among particular groups, such effects may be considerably larger than usual. Second, therefore, we report the results from some of our efforts to discover instances of more powerful self-fulfilling prophecy effects.

A. ACCURACY MORE THAN SELF-FULFILLING PROPHECY

Throughout the 1980s and early 1990s, social psychology abounded with testimonies to the power of expectations to create social reality (e.g., Fiske & Taylor, 1984; Hamilton et al., 1990; Jones, 1986, 1990; Snyder, 1984; see Jussim, 1991, for a review). In contrast, most educational and developmental psychologists argued that expectancy effects were generally minimal, especially in research involving children (Brophy, 1983; Brophy & Good, 1974; Cooper, 1979; Eccles, 1985; Blumenfeld, 1985; Eccles & Wigfield, 1985; West & Anderson, 1976). Evidence from naturalistic studies consistently failed to support the strong claims of social psychologists, and instead confirmed the perspective of the educational and developmental psychologists, rarely uncovering expectancy effects larger than .1 to .2 in terms of standardized regression coefficients (see Jussim, 1991; Jussim & Eccles, 1995, for reviews). Furthermore, research in educational settings has repeatedly shown that teacher expectations predict student achievement mainly because they are accurate (see Brophy, 1983; Jussim, 1991; Jussim & Eccles, 1995, for reviews). Because two of our studies provided some of the clearest evidence of teacher accuracy to date (Jussim, 1989; Jussim & Eccles, 1992), we describe them as follows in some detail.


1. THE DATA

All studies described in this chapter are based on the Michigan Study of Adolescent Life Transitions (MISAL'T), which assessed a variety of social, psychological, demographic, and achievement-related variables in a sample that included more than 200 teachers and 3000 students in the sixth and seventh grades (see Eccles et al., 1989; Midgley, Feldlaufer, & Eccles, 1989; Wigfield, Eccles, Maclver, Reuman, & Midgley, 1991, for more details about this project). A total of about 100 teachers and 1700 students in sixth-grade math classes were the focus of the two studies we summarize here. Both studies tested the hypotheses that 1) teacher expectations early in the year are based on students' previous achievement and motivation and that 2) teacher expectations, student motivation, and students' previous achievement influence students' subsequent achievement (for detailed descriptions of the models and analyses, see Jussim, 1989; Jussim & Eccles, 1992).

Three sixth-grade teacher expectation variables were assessed in early October: perceptions of students' performance, talent, and effort at math. We assumed that teachers inferred students' effort and talent, in part, from their own perceptions of students' performance. Measures included student motivation self-concept of math ability, intrinsic and extrinsic value of math, and self-reports of effort and time spent on math homework. Fall and spring assessments of these motivational variables were included in Jussim (1989); only fall assessments were included in Jussim and Eccles (1992). There were two measures of previous achievement: final marks in fifth-grade math classes and scores on standardized achievement tests taken in late fifth or early sixth grade. There were two outcome measures of achievement: Final grades in sixth-grade math classes and scores on the math section of the Michigan Educational Assessment Program (MEAP), a standardized test administered to students in Michigan early in seventh grade. All measures were reliable and valid (for more detail, see Eccles [Parsons], Adger, & Meece, 1984; Eccles-Parsons, Kaczala, & Meece, 1982; Jussim, 1987, 1989; Jussim & Eccles, 1992, 1989).
to Jussim and Eccles (1992). Although the main analyses were performed using the LISREL VI program, all tests reported below are interpretable as standardized regression coefficients. These two studies were the first to explicitly assess and compare self-fulfilling prophecy, perceptual bias, and accuracy. Both studies assessed models that were more complex versions of the model presented in Figure 1. In brief, we assessed whether teacher perceptions early in the school year predicted changes in achievement (by controlling for previous achievement) over and above changes accounted for by motivation (self-concept of ability, valued placed on math, effort, etc.). Table 1 summarizes the major results from both studies.

Consistent with the self-fulfilling prophecy hypothesis, teacher perceptions of students' math performance in October of the sixth grade significantly related to students' final grades in sixth-grade math (betas = .21, .34) and students' seventh-grade MEAP scores (betas = .10, .15). In Jussim's (1989) study 1 teacher perceptions of talent significantly related to both sixth-grade math grades (beta = .12) and seventh-grade MEAP scores (beta = .17); and 2) teacher perceptions of performance significantly predicted changes in students' self-concept of math ability across the sixth-grade school year (beta = .11).

Results consistent with the perceptual bias hypothesis showed that teacher perceptions of students' effort significantly predicted sixth-grade math grades (betas = .19, .19) to a larger extent than they predicted seventh-grade MEAP scores (betas = .0, .07). Teachers assigned higher grades to students whom they believed had exerted more effort. This hypothetically could have represented accuracy—if teachers were rewarding students who actually were working harder. Instead, however, as the results suggest, teachers simply assumed that higher achievers were working harder. Wherein we found no evidence that the students who received the higher grades actually worked any harder than their peers. In fact, the students who received low grades reported spending more time on homework than the other students (Jussim, 1982). Because effort is difficult to observe directly, we speculated that teachers, perhaps influenced by a belief in a just world (Lerner, 1980) or by the Protestant work ethic (Schuman, Waks, Olson, & Elderidge, 1985; Weber, 1930), simply assumed that "hard work pays off." Therefore, high achievers "must" have been working harder. A consequence, however, is that the academically "rich" (the high achievers) get richer (teachers assign them grades that are even higher than they deserve).

There was both accuracy and inaccuracy in teacher perceptions. Teacher perceptions were largely accurate because they were most strongly linked to appropriate factors in previous grades, standardized test scores, teacher

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Correlation with grades</th>
<th>Effect on grades</th>
<th>Correlation with MEAP scores</th>
<th>Effect on MEAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers' perceptions of students' performance</td>
<td>.54, .71</td>
<td>.57, .64</td>
<td>.52, .83</td>
<td>.65, .73</td>
</tr>
<tr>
<td>Teachers' perceptions of students' talent</td>
<td>.21, .34</td>
<td>.12, .64</td>
<td>.19, .19</td>
<td>.21, .29</td>
</tr>
<tr>
<td>Teachers' perceptions of students' effort</td>
<td>.46, .54</td>
<td>.57, .49</td>
<td>.34, .36</td>
<td>.57, .55</td>
</tr>
<tr>
<td>All three</td>
<td>.10, .15</td>
<td>.17, .12</td>
<td>.06, .07</td>
<td>.33, .37</td>
</tr>
</tbody>
</table>

Effects refer to standardized total effects (see Jussim, 1989, and Jussim & Eccles, 1992, for more detail about the models). The first entry within each column is for Jussim (1989, N = 459), and the second entry is for Jussim & Eccles (1992, N = 1288). All coefficients greater than .05 are significant at p < .05. The differences between correlations and path coefficients is an index of predictive accuracy (see text for explanation). The column titled "All three" reports the multiple correlation of all the teacher perception variables with grades and MEAP scores in the "Correlation" rows and reports the multiple semipartial correlation (controlling for the student background variables) with grades and MEAP scores in the "Effect" rows. (Reprinted with permission from Journal of Personality.)

grade MEAP scores (betas = .0, .07). Teachers assigned higher grades to students whom they believed had exerted more effort. This hypothetically could have represented accuracy—if teachers were rewarding students who actually were working harder. Instead, however, as the results suggest, teachers simply assumed that higher achievers were working harder, whereas we found no evidence that the students who received the higher grades actually worked any harder than their peers. In fact, the students who received low grades reported spending more time on homework than the other students (Jussim, 1982). Because effort is difficult to observe directly, we speculated that teachers, perhaps influenced by a belief in a just world (Lerner, 1980) or by the Protestant work ethic (Schuman, Waks, Olson, & Elderidge, 1985; Weber, 1930), simply assumed that "hard work pays off." Therefore, high achievers "must" have been working harder. A consequence, however, is that the academically "rich" (the high achievers) get richer (teachers assign them grades that are even higher than they deserve).

There was both accuracy and inaccuracy in teacher perceptions. Teacher perceptions were largely accurate because they were most strongly linked to appropriate factors in previous grades, standardized test scores, teacher
perceptions of in-class performance, and student motivation (the multiple correlation of these factors with teacher expectation variables ranged from about 0.6 to 0.5). The results also provide evidence of a small but consistent pattern of zero-order bias in teacher perceptions, but we discuss this issue in detail later in the chapter.

Results from both studies also provided considerable evidence of predictive accuracy. The zero-order correlation between teacher expectations early in the year and student achievement late in the year equals expectancy effects (influences of teacher expectations on student achievement) plus predictive accuracy (teachers basing their expectations on factors that influence student achievement). Therefore, the one index of predictive accuracy is the difference between the zero-order correlation and the size of the expectancy effects (see Jussim, 1989, 1991, 1993; Jussim & Eccles, 1992 for more detailed explanations).

The zero-order correlations of teacher perceptions with seventh-grade MEAP scores ranged from .34 to .57, and the path coefficients ranged from -.07 to .17. The path coefficients relating teacher perceptions to MEAP scores accounted for about 20–30% of the zero-order correlations between initial teacher perceptions and subsequent MEAP scores; the remaining 70–80% represented predictive validity without influence (i.e., accuracy). There was a similar pattern for final grades in sixth-grade math. Zero-order correlations of initial teacher perceptions with year-end grades ranged from .50 to .71. Path coefficients ranged from .04 to .34. The path coefficients relating teacher perceptions to grades accounted for about 30–40% of the zero-order correlations between initial teacher perceptions and subsequent grades; the remaining 60–70% represented accuracy.

Are these results anomalous? Not at all. Research consistently shows that the zero-order correlations between teacher expectations and student achievement generally range from about .4 to .8, and that path coefficients representing effects of teacher expectations on student achievement are generally about .1 to .2 (see Brophy, 1983; Brophy & Good, 1974; Jussim, 1991; Jussim & Eccles, 1995, for reviews).

3. Limitations

The correlational nature of this research leaves open some alternative explanations for the relation between teacher expectancies and student achievement. However, because this study used longitudinal data, reverse causality influences are not possible. Students’ achievement at the end of the sixth grade could not have caused teacher expectations at the beginning of the sixth grade.

The main limitation involves omitted variables. Path coefficients only reflect causal effects when all relevant causes of student achievement have been included in the model. If teacher expectations and student achievement are both caused by a third variable that has been omitted from the model, then the model may yield inflated path coefficients relating teacher expectations to student achievement (a “spurious” relation). Unfortunately, no matter how many potential sources of spuriousness are assessed, it is impossible to know if all such sources have been included.

Although this problem cannot be completely overcome, it can be minimized with the inclusion of extensive control variables. Few naturalistic studies have included more controls (previous achievement test scores and grades, self-concept, and several motivational variables) than we have (Jussim, 1989; Jussim & Eccles, 1992). Thus, these findings provide some of the clearest evidence to date that teacher expectations influence the achievement of some students. Such a conclusion will warrant revision when future research demonstrates empirically that there are important sources of accuracy in teacher perceptions other than those assessed in this study.

It is also important to understand the nature of this limitation. Suppose we omitted important variables that cause both student achievement and teacher expectations. This has a very specific implication—that teachers are even more accurate than we have suggested (i.e., that teacher perceptions predict student achievement less because of their causal influence than because both teacher perceptions and student achievement are based on a third set of variable(s)). Conceptually, of course, this “critique” strengthens our conclusion that teacher expectations predict student achievement more because they are accurate than because they create self-fulfilling prophecies. Nevertheless, it is important to keep these limitations in mind throughout the rest of this chapter. The potential for an omitted variable problem is always present in naturalistic research.
1986; Rosenthal & Jacobson, 1968; Snyder, 1984) that we expected to find effects considerably larger than .1 and .2. Although one of us has taken the lead in arguing that such effects are small (Jussim, 1980, 1990, 1991, 1993; Jussim & Eccles, 1992), this position was based on the empirical data base, not on any preconceived notions that such effects always are, or must be, small. In fact, all the authors of this chapter became interested in self-fulfilling prophecies because of their potential to further understanding social injustice and the construction of social reality. We have, however, challenged researchers to empirically identify naturally occurring conditions under which self-fulfilling prophecy effects are large (Jussim, 1989, 1991, 1993; Jussim & Eccles, 1992). Nonetheless, although expectancy effects may be generally small, we strongly suspected that there were conditions under which expectancy effects were substantially larger. Thus, we embarked on a quest to identify more powerful self-fulfilling prophecies. As the next section shows, we have had some success.

A. STUDENT DEMOGRAPHICS AND SUSCEPTIBILITY TO SELF-FULFILLING PROPHECIES

Understanding the role of demographics in educational and occupational attainment has been a major goal of many of the social sciences for a long time. Research on sources of demographic differences is almost always controversial and highly politicized—regardless of whether it is research arguing for genetic explanations of group difference (e.g., Herrnstein & Murray, 1994) or research castigating that schools oppress girls [American Association of University Women (AAAUW), 1992]. Still, clarion calls and pleas for social psychological research directly addressing issues of race in class, or gender periodically appear in the literature (e.g., Carlson, 1984; Fine & Gordon, 1989; Griffin, 1992). The two phenomena (politicization and lack of research) are probably related. The potential for constructive and outright vilification is so strong that it may deter many scholars from vigorously pursuing research programs into these issues. Yet laboratory researchers often imply or explicitly insist that bias detected by highly artificial procedures or under highly unusual conditions provides insight into biases against African-Americans, women, and people from lower SES backgrounds in naturally occurring social interactions (e.g., Devine, 1989; Greenwald & Banaji, 1995; Von Hippel et al., 1995).

At the minimum there seems to be a broad consensus that issues of race, class, and gender are extremely important. Research on the role of student demographics in moderating teacher expectation effects appears to be particularly important. Such research goes to the heart of the question of whether, how, and how much teacher expectations contribute to social problems and inequality. This alone would be sufficient to justify the exploration of the extent to which student demographics moderate teacher expectation effects.

In addition, however, there were several theoretical arguments leading us to suspect that students from stigmatized groups would be more susceptible to self-fulfilling prophecies than students in general. Abundant evidence suggests that school is often an unfriendly place for many African-American and lower SES students (e.g., Lareau, 1987; Steele, 1992). Although school can be difficult places for both boys and girls, though usually in different ways (e.g., By, 1994; Jussim & Eccles, 1995), math and science classes are often less supportive places for high achieving girls than for high achieving boys (Eccles & Blumenfield, 1985; Eccles & Hoffman, 1984; Eccles-Parns et al., 1982; AAUW, 1992). When school is consistently a difficult place, students may often "disidentify" with achievement by devaluing the importance of school or the particular subjects in which they feel devalued (e.g., Eccles-Parns, 1984; Eccles-Parns et al., 1983; Eccles & Harold, 1992; Jussim, 1986; Meee, Eccles-Parns, Kaczala, Goff, & Putnam, 1982; Steele, 1997). Such responses may render them more readily influenced by teacher expectations in several ways.

When students have a history of negative school experiences find themselves faced with a supportive, encouraging teacher who also insists on high performance, they may feel as if they have caught a breath of fresh air. Such a teacher may inspire previously low achievers to new heights. This perspective may not be as unrealistic as it sounds. In his influential article on Black deindividuation with school, Steele (1992) described academic programs in which previously low-performing students [e.g., some with Scholastic Aptitude Test (SAT) in the 300s] took on difficult heroes-level work and came to outperform their White and Asian classmates. Steele's (1992) description of these programs implies that teachers often engage in behaviors much like those that lead to beneficial self-fulfilling prophecies in the classroom and workplace: They are challenging and supportive (e.g., Brophy & Good, 1974; Cooper, 1979; Eccles & Wigfield, 1987; Ebb, 1984, 1986; Harris & Rosenthal, 1985; Jussim, 1986; Rosenthal, 1989).

However, these same underprivileged students may also be more susceptible to harmful self-fulfilling prophecies. Steele (1992) has articulately argued that students who feel undervalued and "marked" by stigma have fewer defenses against failure. Therefore, even if students do not fail more frequently than students in general, they are more likely to be psychologically devastated by such failures, leading them to "disidentify" with school and achievement. Although Steele's (1992) analysis focused primarily on the plight of African-American students, he speculated that his observations
might also be applicable to girls and students from lower-class social backgrounds.

Although negative teacher expectations are not identical to failure, we speculated that such expectations could readily produce effects analogous to those associated with failure. That is, if students must bear the brunt of inappropriately low teacher expectations, and if they belong to stigmatized groups, their enhanced vulnerability to negative school events may render them more susceptible to self-fulfilling prophecies. Social class and sex (at least in math and science classes) may be at least somewhat similarly stigmatizing. Poor students are frequently seen as inferior to their middle-class peers (Dusen & Joseph, 1983; Rist, 1970), and girls are often viewed as less skilled at math than are boys (Eccles & Hoffman, 1984; Eccles & Jacobs, 1986; Jacobs & Eccles, 1992; Meece et al., 1982; Yee & Eccles, 1988). It is widely believed in our culture that females have less ability in math than males (see Eccles, et al., 1983; Jacobs & Eccles, 1985, 1992). To the extent that females themselves have incorporated this stereotype into their own view of mathematics, they may be especially vulnerable to any behavioral indicators from others that are consistent with the stereotype. For example, Jacobs and Eccles (1985) found that mothers were more likely than fathers to lower their view of their daughters' math ability after being exposed to a media campaign reporting innate gender differences in math ability. These results suggested that females (in this case mothers) are more personally influenced than males by messages consistent with gender-stereotypes.

There is another very different reason to suspect that students from stigmatized groups may be more strongly affected by teacher expectations. Social psychology has a long history of research suggesting that stereotypes of stigmatized groups are often inaccurate (e.g., O. Allport, 1954; Hamilton et al., 1990; Jones, 1986, 1990; Miller & Turnbull, 1986). By definition, the more inaccurate an expectation, the greater its potential to create self-fulfilling prophecies. Therefore, because students from stigmatized groups are perhaps more likely to be viewed inaccurately, they may be more strongly influenced by teachers' expectations.

B. CONCEPTUAL MODEL

Figure 2 presents the conceptual model underlying the following research. The model assumes that student backgrounds (previous grades and test scores, motivation, self-concept, etc.) influence both teacher perceptions and students' future performance outcomes. The model further assumes that teachers' perceptions may also influence student performance outcomes.

Fig. 2. Conceptual model of relationships between teacher perceptions and student achievement.

this is captured by the thick, horizontal arrow. Conceptually, this arrow represents self-fulfilling prophecies. The thick, vertical arrow represents the idea that various proposed moderators may increase or decrease the self-fulfilling influence of teacher expectations on student outcomes.

The short thin arrow represents the potential influence of various aspects of student backgrounds on teacher perceptions. The long thin arrow represents the controls we have included in assessing relationships between teacher perceptions and students' future performance. The relationships represented by these thin arrows are not discussed in this section (although those relationships are precisely the focus of another series of studies reported later).

C. DATA ANALYTIC STRATEGY

Three separate sets of models were estimated: One set focused on student sex, a second on student social class, and a third on student ethnicity. Our analyses first assessed a baseline model, which assumed that the control variables (students' fifth-grade final math grades, scores on standardized tests taken in fifth or early sixth grade, self-concept of math ability, effort spent on math, time spent on math homework, and intrinsic and extrinsic value placed on math) and the three teacher perception variables (performance, talent, and effort) predict fifth-grade final grades and seventh-grade
MEAP scores (see Eccles, 1968; Jussim, 1989; Jussim & Eccles, 1992, for more information about these variables). The results for these baseline models are similar to those summarized earlier from Jussim (1989) and Jussim and Eccles (1992) because these analyses are based on students in the same sample.

Next, we assessed the moderation hypothesis. Specifically, we estimated a new model that included three product terms to the original model, terms representing the product of the student demographic category with each of the three teacher perception variables (performance, talent, and effort). The hypothesis is that a particular student demographic characteristic moderated teacher expectation effects could be confirmed if either the block of three product terms or any of the individual predictors significantly predicted achievement outcomes. However, since the three product terms were highly correlated with each other, testing all three simultaneously could artificially reduce the size and significance of all the product terms (e.g., Gordon, 1968), thereby substantially underestimating the role of any one moderator. Consequently, if at least one of the predictors or the block of three moderators significantly increased the R² at p ≤ .10 level, we examined the individual moderators in three steps.

In step one, we examined a model that included only the product term that most strongly predicted the outcome to the base model. This product term significantly predicted the outcome in each of the analyses that we performed. In step two, we added the other two product terms to the model. If this yielded no significant results (either the R² increment, nor the individual coefficients were significant), the final model included the base model plus the first significant product term. However, if step two yielded significant results (either the R² increment or one of the individual coefficients were significant), we included a third step that essentially repeated step one. In step three, we added the stronger of the remaining two product terms to the final model. In this case, the final model included only the base model plus the two significant product terms. No models ever produced three significant product terms. These procedures reduced underestimation of moderator effects due to collectivity among the product terms. Finally, in order to fully explicate the significant moderator effects, we plotted the predicted relationships separately for the two different demographic groups in each analysis (see, e.g., Judd & McClelland, 1989).

1The three product terms were correlated with each other for two reasons: 1) each teacher perception variable was multiplied by the same potential moderator, and 2) the teacher perception variables themselves were moderately to highly intercorrelated (approximately .5 to .8).

2This procedure is "artificial" because it estimates each individual coefficient after controlling for all other variables in the model. This involves not only the real control variables but also the teacher perception-moderator product term as well—in essence, potentially controlling "out" much of the very moderatorial relationship we are attempting to assess.

Except where noted, the N's for these analyses were 1765 for sex, 1020 to 1050 for social class, and 1609 to 1659 for ethnicity. Variations in sample size reflected different patterns of missing data, primarily with regard to family income and parent education. All analyses reported below were multiple regressions in which the student was the unit of analyses. Because teachers rated all of the students in their classrooms, teacher perceptions are not independent of one another. However, all analyses included classrooms (coded as dummy variables) as predictors of final grades in sixth-grade math and MEAP scores, thereby rendering all other relationships independent of teachers.

D. STUDENT SEX

Were girls more susceptible to self-fulfilling prophecies than were boys? Although student sex did not significantly interact with teacher perceptions to predict MEAP scores, the interaction did significantly predict final sixth-grade math marks. The block of three interaction terms predicted final marks within our significance criterion (R²Sex = .23, p < .001). The only statistically significant product term was Talent × Sex (Talent refers to teacher perceptions of talent). We then reestimated the model using only the one Talent × Sex product term, which significantly (p < .05) predicted grades. The results from this final analysis for predicting grades are presented in Table II.

The regression analysis yielded the following simplified prediction equation:

grades = 10.58 + .23(Talent) + .39(Sex) − .11(Talent × Sex)

We refer to this as a "simplified" prediction equation because it contains only those coefficients and variables directly relevant to understanding how student sex moderates teacher expectation effects. 3

3In creating this simplified prediction equation, we have assumed that all other variables (i.e., other than those in the simplified prediction equation) are at their mean. With this assumption, all other variables yield a constant effect. The constant in the simplified prediction equation equals the constant from the full regression equation plus the product of each variable's coefficient and its mean (each variable except for Talent, Sex, and the Talent × Sex term, which are shown explicitly in the full prediction equation). Consider the following oversimplified example:

If grades = 1 + 10(standardized test scores) + 5(previous grades) + 3.5(Talent) + 2(Sex) + 0.5(Talent × Sex), if the mean standardized test score is 50 (50th percentile), and if the mean grade is 10 (translating letter grades to a numerical scale), then the simplified prediction equation becomes:

grades = 11 + 2(Talent) + 2(Sex) + (Talent × Sex)

The new constant of 11 = 1 + 10(50) + 5(50) + 3.5 + (50 × 0.5) + 10.
In our data, girls were coded as "1" and boys as "2." Therefore, the simplified prediction equation showed that the equations relating teacher perceptions of talent to grades for girls and boys were:

Girls: Grades = 10.97 + .12(Talent)

Boys: Grades = 7.36 + .01(Talent)

These equations were obtained by simply entering the values for student sex into the simplified prediction equations (e.g., Judd & McClelland, 1989).

For example, for girls:

20.58 + .39(sex) + .23(Talent) = 10.97 + .39*(1) + .23(Talent) = 10.97 + .12(Talent)

This equation shows that the slope for girls (.12) is steeper than the slope for boys (.01) and the test of the product term has already shown that this difference is statistically significant. Figure 3 displays the relationship between teacher perceptions and grades separately for boys and girls. It clearly shows that boys' grades are virtually unaffected by teacher perceptions of talent, whereas girls' grades are affected. Even for girls, though, Figure 3 shows the effect to be quite small. The whole range of teacher
perceptions is barely enough to make a difference of one unit in student grades (e.g., B to B+).

Results reported thus far have been unstandardized. Although unstandardized coefficients are often preferable to the standardized coefficients (see Judd & McClelland, 1989; Pedhazur, 1982), standardized coefficients have one major advantage: They easier results from different studies comparable. For example, as discussed previously, effect sizes (in terms of correlation coefficients or standardized regression coefficients) of teacher expectations > student achievement are typically .1 to .3. How do the separate coefficients for boys and girls compare to this general pattern? The standardized coefficient relating teacher grades for girls was .06, and for boys was .14. Thus, even for girls, for whom the effect is strongest, it is still quite small.

E. SOCIAL CLASS

Were students from lower SES backgrounds more vulnerable to self-fulfilling prophecies? Because there were two measures of social class (family income and education), we used multivariate regression in our data analysis approach. Instead of three product terms, six product terms were added to the equation predicting MEAP scores and final grades. Three product terms were created by multiplying each of the three teacher expectation variables by mothers' education; three more terms were created by multiplying each of the three teacher expectation variables by family income.

Although adding these six terms led to a significant F^2 increase ($F_{(12,388)} = 3.97, p < .01$) in the prediction of MEAP scores, none of the coefficients relating the individual product terms to MEAP scores were statistically significant. This reflects collinearity among the product terms. Consequently, we used procedures like those described earlier to identify the most parsimonious subset of the significant predictive product terms:

In this case, the product terms for teacher perceptions of effort*income and teacher perceptions of performance*education, each predicted MEAP scores ($F < .05$). The results from this analysis are displayed in Table III. As with our student sex analyses, we obtained a simplified prediction equation by setting all variables that were not involved in the significant product terms to their mean. This yielded the following simplified prediction equation:

$$ \text{MEAP} = \text{Effort} \times \text{Income} + \text{Performance} \times \text{Education} $$

*Income was coded: 1 = less than $15,000/yr; 2 = $15,000-$30,000/yr; 3 = $30,000-$60,000/yr; 4 = $60,000-$90,000/yr; 5 = more than $90,000/yr. Education was coded: 1 = some high school; 2 = high school; 3 = 2 years of college; 4 = associate's degree; 5 = bachelor's degree. 6 = master's degree; 7 & 8 = doctoral degree.
MEAP = 16.14 + 0.30(ESD) + 1.43(Performance) + .39(Education) + .72(Income) - .33(Education*Income) - 22(Performance*Education),

where ESD refers to teacher perceptions of students' effort and Performance refers to teacher perceptions of students' performance.

Among students whose parents had a lower education (some high school, coded as 2), the unstandardized relationship of teacher perceptions of performance to MEAP scores was .99 (.25, standardized). Among students whose parents had a higher education (having completed college, coded as 6), the unstandardized relation was .11 (.25, standardized). Among students from lower income families (family income of $10,000-$20,000/year, coded as 2), the unstandardized relationship of teacher perceptions of effort to MEAP scores was .04 (.01, standardized). Among students from higher income families (greater than $40,000/year, coded as 1), the unstandardized relationship of teacher perceptions of effort to MEAP scores was actually - .48 (-.15, standardized).

Figure 4 depicts the relationships of teacher perceptions of performance and effort to MEAP scores separately for students from lower and higher SES backgrounds. Students from lower social class backgrounds were dramatically more vulnerable to self-fulfilling prophecies than were their more well-off classmates. As shown in Figure 4, the entire range of teacher perceptions makes a big 4-point difference on the MEAP. In terms of simple percentages, going from 18 to 22 to 0 on the MEAP means going from the 17th percentile to the 58th; going from 21 to 25 means going from the 31st percentile to the 66th; and going from 24 to 28 means going from the 55th percentile to the 99th.

The negative relationship of teacher perceptions of effort to MEAP scores for upper class students completely accounts for the declining slope in Figure 4. Although anomalous, attribution theory does provide one possible reason for why the negative relationship emerged among high SES students. Performance is often assumed to be influenced in a compensatory way by both effort and ability. If one has high ability, then less effort is needed to achieve the same level of performance as that for someone with less ability (Covington & Omissi, 1979). This compensatory relationship may suggest the reason why attributing one’s child’s or one’s students’ success to diligent effort might lead to both lowered ability self-concepts in the child or student and lowered perception of one’s child’s or students’ abilities (Yee & Eccles, 1988). To the extent that a teacher is rating a high SES student as working very hard in his or her class, the teacher may also be conveying the incorrect message that the student has low ability. This message could then undermine the student’s motivation or increase the student’s anxiety such that the student performs more poorly in a standardized testing situation.

Alternatively, the teacher’s view that the student is working hard may really mean that the student does work harder than other students to compensate for lower ability. If so, this lower ability level could explain why the student does not do as well as his or her peers in a closed standardized testing situation, in which there is insufficient time for greater effort to compensate for lower ability in determining final performance level.

A similar pattern was obtained for final grades. Although the R² increment associated with adding all six product terms approached significance (F(6, 590) = 1.71, p = .12) only when entered alone, the teacher perceptions of performance by income product term did significantly predict grades (p < .01). The final model, then, included only this one product term. These results are summarized in Table 4. This analysis yielded the following simplified prediction equation:

Grades = 8.54 + .88(Performance) + .33(Income) - .29 (Performance*Income),

where Performance refers to teacher perceptions of performance.
### TABLE 10.2: From Measuring Analyse Final Marks Predicting Grades

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>First-year</th>
<th>English</th>
<th>Math</th>
<th>Science</th>
<th>Socioeconomic Status</th>
<th>EF</th>
<th>Perception of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall N</td>
<td>15,901</td>
<td>12,931</td>
<td>13,901</td>
<td>13,901</td>
<td>13,901</td>
<td>13,901</td>
<td>13,901</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>2,253</td>
<td>2,233</td>
<td>2,230</td>
<td>2,240</td>
<td>2,232</td>
<td>2,240</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>13,648</td>
<td>13,708</td>
<td>13,671</td>
<td>13,661</td>
<td>13,669</td>
<td>13,661</td>
</tr>
<tr>
<td>College</td>
<td>Main</td>
<td>12,738</td>
<td>12,714</td>
<td>12,718</td>
<td>12,722</td>
<td>12,726</td>
<td>12,728</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>12,738</td>
<td>12,714</td>
<td>12,718</td>
<td>12,722</td>
<td>12,726</td>
<td>12,728</td>
</tr>
<tr>
<td>Income</td>
<td>Low</td>
<td>7,970</td>
<td>7,948</td>
<td>7,952</td>
<td>7,956</td>
<td>7,958</td>
<td>7,956</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>7,768</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
</tr>
<tr>
<td>Race</td>
<td>Black</td>
<td>7,970</td>
<td>7,948</td>
<td>7,952</td>
<td>7,956</td>
<td>7,958</td>
<td>7,956</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>7,768</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
<td>7,766</td>
</tr>
</tbody>
</table>

**Table Notes:**
- **Dependent variable:** Final marks predicting grades.
- **First-year:** Freshman year.
- **English:** English language.
- **Math:** Mathematics.
- **Science:** Science subjects.
- **Socioeconomic Status:** Socioeconomic status.
- **EF:** Effort.
- **Perception of Performance:** Students' perception of their performance.
- **Overall N:** Total number of students.
- **Gender:** Male and Female.
- **College:** Main and Subtotal.
- **Income:** Low and High.
- **Race:** Black and White.

---

**Accuracy and the Self-Fulfilling Prophecy**

![Fig. 5. Teacher expectations influence the grades of students from lower SES backgrounds more strongly than they influence the grades of students from higher SES backgrounds. Lower income refers to an income of $15,000-$20,000. Higher income refers to an income of greater than $40,000.](image)

The relationship of teacher perceptions of performance to grades for students from low income backgrounds (family income of $10,000-$20,000) was .56 (unstandardized) and .51 (standardized). The relationship of teacher perceptions of performance to grades for students from higher income backgrounds was .43 (unstandardized) and .19 (standardized). These results, displayed in Figure 5, clearly show that teacher perceptions influenced the grades of lower income students more strongly than they influenced the grades of upper income students. Comparing these results to those displayed in Figure 4, however, shows that the difference is less dramatic for grades than for MEAP scores. Across the entire range of grades, the teacher expectation effect made a difference of about 2 grade levels for upper income students (e.g., C+ to B) and three grade levels for lower income students (e.g., C+ to A+).

**F: ETHNICITY**

Were African-American students more susceptible to self-fulfilling prophecies than White students? To answer this question, three ethnic-

---

*Grades were coded on a scale ranging from 3 (F) to 16 (A+), with each unit representing the same grade level: grade 4 was F; grade 5 was D; grade 6 was C; grade 7 was B; grade 8 was A; grade 9 was A+.\]
ity by teacher perception product terms were added to the equations predicting MEAP scores and final grades. The analysis on MEAP scores was based on the 1536 White students and 72 African-American students who had valid data on all variables. Results are summarized in Table V.

The final analysis yielded a significant ethnicity by teacher perception of performance product term ($\beta = -.14; t = 2.79, p < .01$). The simplified prediction equation was:

$$\text{MEAP} = -24.18 - 3.24(\text{ethnicity}) - .35(\text{Performance})$$
$$+ .94(\text{Performance} \times \text{ethnicity})$$

Ethnicity was coded as 1 for White students and 2 for African-American students. The relationship of teacher perceptions of performance to MEAP scores was .59 (unstandardized) and .14 (standardized) for White students and .53 (unstandardized) and .37 (standardized) for African-American students. Figure 6 displays these relations and clearly shows the dramatically greater expectancy effects for African-American students. For White students, the entire range of teacher perceptions of performance makes about a 2.5-point difference score on the MEAP, whereas for African-American students, that range makes about a 6-point difference.

To make this finding more concrete, consider a two- or three-point difference in MEAP scores (the magnitude of the largest changes associated with teacher perceptions among White students). Scores going from 19 to 22 would mean going from the 31st to the 38th percentile, and going from 24 to 26 would mean going from the 55th to the 78th percentile. Now consider a six-point difference (the magnitude of the largest possible changes associated with teacher perceptions among teacher perceptions with expectations for African-American students). Going from 17 to 23 would mean going from the 14th to the 45th percentile, and going from 21 to 27 would mean going from the 31st to the 89th percentile.

Analyses examining predictors of final grades, which included 76 African-American students and 1587 White students, showed a similar pattern. Adding the three product terms significantly increased the $R^2$ increment ($F(3,1552) = 6.91, p < .001$). In the final model, however, only the ethnic-

---

1. A dotted line indicates the model.
2. Ethnicity by teacher perception of effort product term was also statistically significant ($p < .05$). However, 1) its coefficient was negative, 2) the unstandardized coefficient for the ethnicity by performance interaction doubled, going from about .0 to the $R^2$ in the main test to .1, and 3) the coefficient for ethnicity by effort was still almost identical to the effect reported in the main text. Because the analysis with only one product term is simpler to present and interpret, and provides essentially the same information as the analysis with two product terms, the main text reports the analysis with the one product term.
ity by teacher perceptions of performance products was significant (adding the other two terms, individually or together, yielded no additional significant individual coefficients and no significant $R^2$ increase). These results are summarized in Table VI. The simplified prediction equation was:

$$\text{Grades} = 13.05 - 0.11(\text{Performance}) - 2.11(\text{Ethnicity}) + 0.64(\text{Performance} \times \text{Ethnicity})$$

The results (displayed in Figure 7), are consistent with the prediction that teacher expectation effects are stronger among African-American students than among White students. For White students, the relationship of teacher expectations of performance to final grades was $r = 0.66$ (unstandardized) and $r = 0.70$ (standardized), and for African-American students, the relationship was $r = 0.18$ (unstandardized) and $r = 0.23$ (standardized). Figure 7 shows that, among other variables being held equal, going from the lowest to the highest teacher perception produced a 4-unit change in grades (e.g., going from C- to B+) among African-American students, but only a 2-unit change in grades (e.g., going from C+ to B) among White students.

We were concerned about two limitations to our ethnicity study. First, the sample of African-American students was quite small. Because attention came mainly from students moving in and out of the district during the 3-year span covered by the analysis (fifth through seventh grades), we were able to increase the sample size by omitting some seventh-grade data. We
African-American Students

White Students

Grades

-8

-7

-6

-5

-4

-3

-2

-1

0

1

2

3

4

5

Teacher Perceptions of Performance

Fig. 7. Teacher expectations influence the grades of African-American students more strongly than they influence the standardized test scores of White students.

G. UNCONFONDING THE EFFECTS OF ETHNICITY AND SOCIAL CLASS

A common problem in research on ethnicity and social class is assessing their separate roles in social phenomena. This is a problem because, on the average, Whiteness from higher SES backgrounds than many minorities. However, this was not a major problem in the current study. Ethnicity was uncorrelated with parental education ($r = -0.01$, ns), and only weakly related to income ($r = -0.18$, $p < .001$). Nonetheless, several additional analyses further probed this issue. First, we reran the main social class analyses (see Tables III & IV) excluding all African-American students. The results were virtually identical to those reported in Tables III and IV, indicating that results regarding social class do not derive primarily from the African-American students.

Second, despite some important limitations, we ran an additional set of analyses adding both the ethnicity and SES product terms (together) to the model predicting grades and MEAP scores. One limitation is that, because family income and education information was available for only slightly more than half of the sample, the number of African-American students was reduced even further—to 27. Nonetheless, the results for the analysis predicting MEAP scores were striking: All three two-way product terms (involving education, income, and ethnicity) significantly ($p < .05$) predicted MEAP scores, and the regression coefficients were nearly identical to those reported in Tables III and IV.

Results for the model predicting grades were similar but weaker. Here, colorateness was a nuisance. Adding the performance by income and performance by ethnicity terms to the base model significantly increased the $R^2$ increment ($F(2, 897) = 4.35, p < .05$). However, neither product term, individually, reached statistical significance. The performance by income coefficient was similar to the term shown in Table IV (—.09 in the table versus —.07 in the current analysis, $p < .05$), but the ethnicity by performance coefficient was lower (.64 in the table versus .32 in the current analysis, $p = .12$).

These results, despite their limitations, were similar to those obtained in the analyses that addressed ethnicity and social class separately. It seems,
therefore, that the higher expectancy effects among lower SES students are not due to the greater proportion of African-American students from lower SES backgrounds. It is also unlikely that the greater expectancy effects among African-American students are due to their lower SES backgrounds.

It MULTIPLE VULNERABILITIES

These analyses, however, raise a more general question: Are students who are members of more than one vulnerable group more susceptible to expectancy effects than other students? For example, are lower SES African-American students more susceptible? Are lower SES girls also doubly vulnerable? Are African-American girls and lower SES African-American girls also especially vulnerable? The results just reported suggest that the greater vulnerability of lower SES students is largely independent of the greater vulnerability of African-American students. Put differently, they suggest that these effects are additive, which implies that, because SES and ethnicity are both powerful moderators, lower class African-American students would be the most vulnerable of all to teacher expectation effects. Hypothetically, these questions could be answered by adding the requisite three-way product terms. Unfortunately, however, because African-American students with parental SES information are relatively few in number, the results from a model including three-way product terms combining ethnicity, SES, and teacher perceptions would not have been meaningful. Because we are presently unable to include these three-way product terms, direct assessment of this question must await future research.

We did, however, assess whether lower SES girls and African-American girls were particularly susceptible by assessing models including three-way product terms (between student sex, teacher perceptions, and either SES or ethnicity). Because these models are very complex (base model, plus two-way product terms for all combinations of sex by the three teacher perception variables by either the two SES variables or by ethnicity, plus three-way product terms) we can only summarize our analysis here.

There were four main analyses (Two outcomes) by two separate tests of three-way interactions: teacher perceptions by student sex by student SES predicting grades and MEAP; and teacher perceptions by student sex by student ethnicity predicting grades and MEAP). Each analysis was performed in two steps. The first step included the base model, plus all lower-order two-way product terms. Previously obtained significant two-way product terms remained significant in all of these models. In the second step, we added the three-way product terms. For three of the four models, none of the three-way product terms significantly predicted student outcomes. Even the fourth did not qualify our previous conclusions. It showed that, for MEAP scores, lower SES girls were slightly more susceptible to expectancy effects than were other groups.

These results mean that lower SES girls and African-American girls are more susceptible to expectancy effects than are students who belong to only one vulnerable group (i.e., the two interaction effects were generally additive). These results are consistent with our general conditions that girls are slightly more susceptible than boys, that lower SES students are considerably more susceptible than upper SES students, and that being a lower SES or African-American girl is a double vulnerability.

V. Stereotypes and Self-Fulfilling Prophecies

Teacher stereotypes would appear to be a likely explanation for why expectancy effects are more powerful among students from stigmatized or disadvantaged social groups. Both the evidence regarding the self-fulfilling effects of social stereotypes and the limitations to that evidence are discussed next.

Many social psychological perspectives and reviews claim or assume that stereotypes are often inaccurate and could likely lead to self-fulfilling prophecies (Hamilton et al. 1990; Jacobs & Eccles, 1992; Miller & Turnbull, 1986; Snyder, 1964; von Hippel et al., 1995). The role of stereotypes in creating self-fulfilling prophecies that contribute to inequalities between ethnic and socio-economic groups and between the sexes may seem "obvious." The dominant group (White men) hold negative stereotypes about other groups. White men treat members of these groups less favorably than they treat other White men, so that members of these other groups receive lower quality education (and lower paying jobs, too).

Unquestionably, this sequence sometimes occurs. Although the claim that social stereotypes are self-fulfilling appears straightforward, it is considerably more complex than it seems for several empirical and conceptual reasons. Although any stereotype, hypothetically, may be self-fulfilling, most research to date has focused on four particular stereotypes: ethnicity, social class, gender, and physical attractiveness. This research literature is reviewed next, after which we address basic theoretical issues involved in understanding the extent to which stereotypes create social injustices through self-fulfilling prophecies.

A. ETHNIC STEREOTYPES

We are aware of only one study that comes close to documenting self-fulfilling prophecies produced by an ethnic stereotype. In the first class