

Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies

Lee Jussim

*Department of Psychology
Rutgers University*

Kent D. Harber

*Department of Psychology
Rutgers University at Newark*

This article shows that 35 years of empirical research on teacher expectations justifies the following conclusions: (a) Self-fulfilling prophecies in the classroom do occur, but these effects are typically small, they do not accumulate greatly across perceivers or over time, and they may be more likely to dissipate than accumulate; (b) powerful self-fulfilling prophecies may selectively occur among students from stigmatized social groups; (c) whether self-fulfilling prophecies affect intelligence, and whether they in general do more harm than good, remains unclear, and (d) teacher expectations may predict student outcomes more because these expectations are accurate than because they are self-fulfilling. Implications for future research, the role of self-fulfilling prophecies in social problems, and perspectives emphasizing the power of erroneous beliefs to create social reality are discussed.

Teacher expectations. The term has been known to inspire righteous indignation for teachers' supposed role in creating inequalities. The key reason is the self-fulfilling prophecy (Merton, 1948)—erroneous teacher expectations may lead students to perform at levels consistent with those expectations (Brophy & Good, 1974; Rosenthal & Jacobson, 1968).

It is not clear, however, that the evidence justifies condemnations of teachers for their supposed role in creating injustices. Other researchers, equally righteous, condemn some teacher expectation research as appallingly flawed and have expressed dismay at how it has been misinterpreted and has captured the popular imagination (Elashoff & Snow, 1971; Snow, 1995; Wineburg, 1987). This review conveys why 35 years of research on teacher expectations leads to conclusions that fall between these extremes. Although some specific teacher expectation studies may have suffered flaws sufficiently serious to threaten their conclusions, the abundant naturalistic and experimental evidence shows that teacher expectations clearly do influence students—at least sometimes.

Furthermore, research has continued to provide new data bearing on several classic and highly speculated-on aspects of teacher expectations and self-fulfilling prophecies. Although pieces of this literature have been reviewed elsewhere (Jussim, Smith, Madon, & Palumbo, 1998; Spitz, 1999), and although social psychological reviews of expectancies often at least mention teacher expectations (Olson, Roese, & Zanna, 1996), no broad review summarizing and synthesizing what is known and unknown about teacher expectations and self-fulfilling prophecies has appeared in about 20 years (since Brophy, 1983). Many prior reviews and meta-analyses of teacher expectations have focused on (a) particular scientists' programs of research (Jussim et al., 1996; Weinstein & McKown, 1998); (b) arguments for or against the reality of self-fulfilling prophecies in general or in specific types of studies or contexts (Raudenbush, 1984, 1994; Rosenthal, 1974, 1994; Snow, 1995; Spitz, 1999; Wineburg, 1987); or (c) documentation of the processes by which self-fulfilling prophecies in the classroom occur (Brophy, 1983; Brophy & Good, 1974; Harris & Rosenthal, 1985; Jussim, 1986).

Reviews that have addressed general self-fulfilling prophecies, by their nature, could not focus exclusively, or even primarily, on teacher expectation effects (Darley & Fazio, 1980; Jones, 1986, 1990; Jussim, 1991; Miller & Turnbull, 1986; Olson et al., 1996; Snyder & Stukas, 1998). Therefore, a review highlighting how empirical research has addressed several core

We sincerely thank David Funder for comments on earlier drafts of this article. This article is based, in part, on sections from a book by Lee Jussim, currently in progress and tentatively titled *Social beliefs and social realities*.

Requests for reprints should be sent to Lee Jussim, Department of Psychology, Rutgers University, Piscataway, NJ 08544. E-mail: jussim@rci.rutgers.edu

and controversial questions involving teacher expectations might provide a valuable contribution to enduring and significant theoretical questions, such as the extent to which social perception is accurate versus creates social reality, and to applied issues, such as whether teacher expectation effects primarily help create versus alleviate social problems.

The Six Questions on Which This Review Is Focused

This review is framed around six questions that capture many of the central issues addressed by teacher expectation research:

1. What did the early teacher expectation research show?
2. Do teacher expectations influence student intelligence?
3. How powerful is the typical self-fulfilling prophecy in the classroom?
4. How accurate is the typical teacher expectation?
5. Do negative teacher expectations harm students more than positive teacher expectations help students?
6. Do teacher expectation effects accumulate across different teachers and over time?

These six questions were selected for several reasons. In addition to providing a framework for reviewing some of the major themes of research on teacher expectations, each question also reflects a modern controversy. Although Questions 1 through 4 have been discussed in many prior reviews (Brophy, 1983; Rosenthal, 1974; Snow, 1995; Spitz, 1999; Wineburg, 1987), none have synthesized research addressing all four questions in an attempt to reach a set of integrated conclusions. Moreover, the answers to those four questions remain controversial today, given that one can find recent literature suggesting or explicitly espousing diametrically opposed conclusions regarding each question. Furthermore, there are no comprehensive reviews addressing whether expectations are generally harmful (Question 5) and whether expectations accumulate (Question 6), even though speculative claims emphasizing the negative effects of self-fulfilling prophecies, and their power to accumulate, are commonplace. This review, therefore, not only highlights how the accumulated evidence bears on these controversies, but shows how the implications of that research can be integrated into a relatively small number of straightforward conclusions.

Each question, furthermore, goes to the heart of claims emphasizing the potential role of teacher expectations and self-fulfilling prophecies in social problems. This potential appears to be a central reason for

some of the heat in the controversies that have enveloped this area of research from the outset. Specifically, the social problems view of teacher expectations suggests that the answer to the six questions are that self-fulfilling prophecies are real, widespread, and powerful; they can have profound effects, not just on achievement, but on intelligence; they are frequently inaccurate; negative expectancy effects are stronger than positive ones; and small effects accumulate over time (for reviews emphasizing the role of teacher expectations or self-fulfilling prophecies in social problems, see, e.g., Claire & Fiske, 1998; Darley & Fazio, 1980; Fiske & S. Taylor, 1991; Hamilton, Sherman, & Ruvolo, 1990; Jones, 1986, 1990; Schultz & Oskamp, 2000). The validity of this social problems perspective on teacher expectations will be evaluated throughout this review.

Sticking Close to the Evidence

Because this area has been fraught with controversy, it is particularly important for the conclusions reached in any review to stick close to the empirical evidence. In this review, this means two things. First, sticking close to the empirical evidence means that one major criterion for reaching broad and general conclusions will involve relying on the actual results of the relevant studies, rather than on a consensus of scholarly opinion. Although such consensus usually corresponds well with empirical evidence, some disconnects between consensus and evidence are highlighted in this review. Meta-analyses, sometimes of hundreds of studies, have been performed and provide clear information about the average size of the effect, as well as about other points addressed in this article. When such meta-analytic results exist, and when they conflict with claims made in prior narrative reviews, this review will base conclusions regarding effect sizes on the meta-analyses, not on the prior narrative reviews (to use the same metric for experiments, meta-analyses, and naturalistic studies, except where otherwise stated, all effect sizes in this paper are either r 's [correlations] or, in the case of naturalistic studies using structural equation techniques, standardized regression coefficients). This review will also identify potential reasons for the disconnect between the meta-analyses and narrative reviews.

Second, this approach means that, when interpreting a specific study (as will be necessary when only a handful of studies has addressed some issue), the conclusions reached here will focus far more on the actual empirical results provided by that study than on either the original authors' conclusions or on how that study has frequently been interpreted by others. When there is ambiguity in interpreting the research bearing on any

of the six questions, alternative views and interpretations will be presented and critically evaluated.

What This Review Does Not Address

This article does not address the interpersonal processes that create or limit self-fulfilling prophecies, or the evidence regarding bases of teacher expectations, because such issues have been addressed at length previously (Brophy, 1983; Dusek & Joseph, 1983; Harris & Rosenthal, 1985; Jussim, 1986). Because self-fulfilling prophecies require changes in students' actual achievement, and not merely changes in teacher perceptions of student achievement, this review does not address the extent to which expectations bias teachers' judgments of students without changing students' actual achievement (see Jussim, 1989; Williams, 1976, for such research). Furthermore, because this is not a general review of all interpersonal expectations, only empirical studies specifically addressing whether teacher expectations create self-fulfilling prophecies are included here.

Last, although this article does address the accuracy of teacher expectations, it does not address the phenomenology of accuracy. Scientifically, we define teacher expectation accuracy as predictive validity without (self-fulfilling) influence. This definition may differ from teachers' subjective experience of "accuracy." Consider a teacher who successfully predicted a student's performance. Even if that successful prediction derived entirely from a self-fulfilling prophecy, subjectively, the teacher might experience this as "accuracy." Scientifically, however, it is not accuracy at all. Whether teachers do or do not see accuracy in the same manner as described here is an interesting issue, but beyond the scope of this review.

Resilient and Recurring Controversies: What Did the Early Teacher Expectation Research Show?

One might wonder why it is necessary to review research that is over 30 years old and has been reviewed amply elsewhere. First, Rosenthal and Jacobson's landmark *Pygmalion in the Classroom* study (1968) is still regularly cited in support of conclusions that their data did not actually support. Second, modern discussions of teacher expectations draw on this literature to reach conclusions that are all over the map, ranging from emphasizing their power to influence students (Gilbert, 1995; Schultz & Oskamp, 2000), to suggesting that such effects are minimal (Snow, 1995; Spitz, 1999), to denying their existence altogether (Roth, 1995; Rowe, 1995). Thus, in the spirit of "sticking close to the data," the next section first reviews the

original Rosenthal and Jacobson (1968) study and then revisits some of the early follow-up research that assessed the reality of teacher expectation effects. This review identifies what that early research did and did not show and, by extension, what sort of conclusions it does and does not justify.

Pygmalion: How It Was Conducted

Rosenthal and Jacobson (1968) performed a simple and elegant study. They administered the TOGA, a nonverbal intelligence test, to all of the children in Jacobson's elementary school (kindergarten through fifth grade). However, they did not tell the teachers that this was an intelligence test. Instead, special covers labeled it as a "A Test of Inflected Acquisition," which, an information sheet explained, was a new test being developed at Harvard for identifying children likely to "bloom"—to show a sudden and dramatic intellectual spurt over the upcoming school year. Rosenthal and Jacobson (1968) then informed each teacher which of their students had been identified as potential "late bloomers." These late bloomers (about 20% of the total in the school), however, were actually selected at random. As Rosenthal and Jacobson (1968, p. 70) put it, "The difference between the children earmarked for intellectual growth and the undesignated control children was in the mind of the teacher." They then administered the TOGA intelligence test again 1 year later and 2 years later.

Results: The Oversimplified Version

Teacher expectations created a self-fulfilling prophecy. One year later, the late bloomers gained more IQ points than did the control students. Even 2 years later, the bloomers' gains still exceeded those of the control students. Although the only initial systematic difference between bloomers and controls was in the teachers' minds, the late bloomers actually showed IQ gains relative to controls. The teachers' false expectations had become true. Rosenthal and Jacobson's (1968) results also showed that the more the control children gained in IQ, the less well-adjusted, interesting, and affectionate they were seen by their teachers. Teachers seemed actively hostile toward the students showing unexpected intellectual growth.

When described in this manner, these results seem dramatic. Inaccurate high teacher expectations provided an undue advantage to some students. And, when children unexpectedly exceeded teachers' expectations, rather than leading to support and reinforcement, this seemed to trigger oppressive teacher responses toward those students. This was a powerful combination and it seemed to explain how teachers' expectations—and, by extension, the expectations of managers, college admissions personnel, health professionals, and so

on—could be a major contributor to the social inequalities associated with race, sex, and social class.

Results: The Messier, More Complicated, and Truer Version

There is nothing false in the aforementioned, oversimplified summary of Rosenthal and Jacobson, 1968. It is true, and to this day, the study is often described in this manner (Fiske & S. Taylor, 1991; Gilbert, 1995; Myers, 1999; Schultz & Oskamp, 2000). Nonetheless, Rosenthal and Jacobson's (1968) pattern of results was not quite as straightforward as the summary suggests.

One complication was that, on average, both groups of children—late bloomers and controls—showed dramatic IQ gains over the next year. On average, the late bloomers gained about 12 points and the controls about 8 points. This is important for at least two reasons. First, in this study, there was no IQ evidence of teachers harming or oppressing students. Most students gained in IQ, regardless of experimental condition. And the control group's average gain of 8 points is quite dramatic—it is about half of a standard deviation on a typical IQ test. Although the study's results did not preclude the possibility of teacher expectations actively harming students, there was no IQ evidence in this study indicating that such harm actually occurred.

Second, although the across-the-board IQ increases could be described as “dramatic,” the differences between the gains of the late bloomers and the controls were not so dramatic. Averaging across all grade levels, that difference was about 4 points. This difference was statistically significant—but it would be difficult to characterize a 4-IQ-point difference as a “dramatic” effect.

Other ways to consider the size of the effect also yield a picture of a less than dramatic result. The difference between the experimental and control conditions corresponded to an effect size of .30 (difference between the experimental and control group in standard deviation units). Typically, effect sizes of .30 or less are considered small (Cohen, 1988). Or, we could simply correlate the manipulation with IQ scores. That correlation is $r = .15$ (Rosenthal, 1985). The size of the difference between bloomers and controls was something less than dramatic.

There was, however, some evidence of dramatic effects. In the first grade, the bloomer's out-gained the control students by about 15 IQ points; in second grade the difference was about 10 points. In both grades, the control students gained IQ points—but such gains were not even close to those gained by the bloomers.

But the story again becomes more complicated. There was no difference between third-grade bloomers and controls. In fourth grade, bloomers gained more than controls, but the difference was not statistically significant. In fifth and sixth grade, bloomers actually

gained fewer IQ points than did controls, but this difference was not statistically significant either. Thus, the overall effect averaged across all six grades derived almost entirely from the effects in first and second grade. A theoretically coherent and compelling account, however, might still be maintained by arguing that young children were more susceptible to teacher expectation effects. The ability of this explanation to account for Rosenthal and Jacobson's data, however, is more apparent than real.

After 2 years, the oldest children (then in sixth grade) showed the largest differences between bloomers and controls. If there was much greater “susceptibility” among younger children, it did not last very long. And what mechanism could explain why, among the older children, there was a complete absence of a teacher expectation effect in Year 1 but the largest effects obtained in Year 2? (Because there remains no empirical evidence supporting any such explanation, and because no follow-up research has replicated this pattern, it will not be discussed further.) Such odd patterns considerably muddied the interpretive waters surrounding the study.

The Extreme Reactions to this Study

Enthusiastic acceptance of the study by the general intellectual public and many social scientists.

The Pygmalion study hit a sensitive social and political nerve (see reviews by Spitz, 1999; Wineburg, 1987). It has frequently been cited in support of the following line of argument: (a) erroneous social stereotypes are a common source of expectations; (b) teacher expectations are self-fulfilling; so that (c) teacher expectations are potentially a powerful force in the creation of social inequalities and injustices (Claire & Fiske, 1998; Hofer, 1994; Jones, 1986, 1990; Rist, 1970; Schultz & Oskamp, 2000; M. C. Taylor, 1992; Weinstein & McKown, 1998, for variations on this line of argument). Especially if this self-fulfilling process occurs, not only in elementary school classrooms, but in colleges, in the workplace, in government, and so on, the phenomenon is capable of accounting for long-term entrenchment of social inequalities. Rosenthal and Jacobson (1968) may have so captured the imagination of the intellectual public and many social scientists, at least in part, because its message was clear and simple and it seemed to provide scientific credibility and strong rhetorical ammunition for pundits, policymakers, social activists, and reformers.

The storm of criticism. Not everyone, however, greeted Pygmalion uncritically. Among some researchers studying educational psychology and intelligence, the study generated a storm of criticism (Jensen, 1969; Thorndike, 1968). Two of these researchers

(Elashoff & Snow, 1971) wrote an entire book critiquing the Pygmalion study. Consider the following, from Snow's (1969) critique of Pygmalion that appeared in *Contemporary Psychology* (p. 197):

The study suffers from serious measurement problems and inadequate data analysis. Its reporting, furthermore, appears to violate the spirit of Rosenthal's own earlier admonitions to experimenters and stands as a casebook example of many of Darrell Huff's (How to Lie with Statistics. New York: Norton, 1954) admonitions to data analysts.

Many of the complaints leveled against the original Rosenthal and Jacobson (1968) study, however, were more flawed than the study itself. One such charge was that the measure of IQ was unreliable (e.g., Roth, 1995; Thorndike, 1968) apparently in an attempt to suggest that any results developed using such a measure were meaningless. In fact, however, lack of reliability in a measure makes it harder to find differences between groups. Therefore, finding differences between groups with a measure low in reliability attests to the power of those differences.

Taking Pygmalion Results at Face Value

Both reactions to the original Pygmalion study—uncritical acceptance and overgeneralization on one hand; vilifying criticism on the other—are probably too extreme. The study has imperfections (as do all studies)—but whether any were so lethal as to invalidate its basic findings remains unclear. A later section of this article addresses in detail some of the strongest evidence against the study's validity. However, even if one takes Rosenthal and Jacobson's results entirely at face value, the justifiable conclusions are considerably more modest than suggested by the overly dramatic manner in which the study has frequently been portrayed.

This section is organized around several questions—questions to which wrong answers have often seemed obvious, or at least implied, in many discussions of the original Pygmalion study.

1. Were teacher expectations typically inaccurate? This was not assessed. Therefore, the Pygmalion study provided no information about the typical accuracy or inaccuracy of teacher expectations.

2. Did demographic-based stereotypes unduly bias expectations and perceptions? Rosenthal and Jacobson (1968) did not assess the extent to which student demographics or social stereotypes influenced teacher expectations. Therefore, the study provided no data bearing on the issue of whether stereotypes bias teacher expectations.

3. Were self-fulfilling prophecies typically powerful and pervasive? They were clearly not typically powerful. The overall effect size equaled a correlation of $r = .15$. The mean difference in IQ gain scores between late bloomers and controls was 4 points. Nor were self-fulfilling effects pervasive. Significant teacher expectation effects only occurred in two of six grades (in Year 1) and in one of five grades in Year 2. This means that self-fulfilling prophecies did not occur in 8 of 11 grades examined.

4. Were powerful expectancy effects ever found? Yes. The results in first and second grade in year one (15- and 10-point bloomer-control differences) were quite large.

5. Were self-fulfilling prophecies harmful? Rosenthal and Jacobson (1968) only manipulated positive expectations. They showed that inaccurate positive expectations could be self-fulfilling. It would have been unethical to instill inaccurate negative expectations. Therefore, they did not assess whether inaccurate negative expectations undermine student IQ or achievement.

The Scientific Contribution of Rosenthal and Jacobson (1968)

For all the drama and controversy, the study's actual findings ranged from nil (if one believes the critics) to quite modest, if taken at face value. This is clearly a case, however, where a study's contribution involved more than its specific results. Rosenthal and Jacobson's (1968) study opened up new areas of research in education and psychology (Brophy, 1983; Brophy & Good, 1974; Miller & Turnbull, 1986; Snyder, 1984).

Nonetheless, given the controversy surrounding the study's actual results, the first order of business for many researchers was to evaluate the validity of the basic teacher expectation/self-fulfilling prophecy phenomenon. That research is summarized next.

Are Self-Fulfilling Prophecies Real?

Rosenthal and his colleagues, and others, attempted many replications of the original Pygmalion study (see reviews by Brophy & Good, 1974; Rosenthal, 1974; Spitz, 1999). Because of the methodological criticisms of the first Pygmalion study, many of the early replications focused not on the general question of whether teacher expectations can be self-fulfilling, but on precise (or narrow) attempts to determine whether experimentally induced erroneous teacher expectations had self-fulfilling effects on student IQ and achievement.

Even these studies, however, evoked controversy. Consistently, only slightly over one third demonstrated a statistically significant expectancy effect (Brophy, 1983; Rosenthal & Rubin, 1978). This pattern seemed to resolve nothing. It was often interpreted by the crit-

ics as demonstrating that the phenomenon did not exist because support was unreliable. Proponents interpreted this result as demonstrating the existence of self-fulfilling prophecies because, if only chance differences were occurring, replications would only succeed about 5% of the time.

Resolution

The Pygmalion controversy was to have an effect that went well beyond self-fulfilling prophecies. In his attempt to refute critics, Rosenthal became one of the pioneers in development of meta-analysis (Harris, 1991). Rosenthal and Rubin's (1978) meta-analysis of the first 345 experiments on interpersonal expectancy effects conclusively demonstrated the existence of self-fulfilling prophecies. The 345 studies were divided into eight categories. *Z* scores representing the combined expectancy effect in all studies in each category were computed. The median of the eight combined *Z* scores was 6.62, indicating that the self-fulfilling prophecy was real.

Still Controversial After All These Years: IQ, Accuracy, Power, Positivity/Negativity and Accumulation

After a decade of heated debate culminating in the resolution presented in Rosenthal and Rubin's (1978) meta-analysis, one might wonder how there could be anything controversial left to discuss. Nonetheless, different researchers have sometimes reached diametrically opposed conclusions regarding the power and patterns of self-fulfilling prophecies, which have occasionally blossomed into heated controversies. The principal issues of contention reflected in the modern literature are whether there is any self-fulfilling effect on intelligence; the accuracy and power of teacher expectations; the relative effects of positive versus negative expectations; and the extent to which self-fulfilling prophecies accumulate versus dissipate over time. These issues are discussed next.

Are Self-Fulfilling Prophecy Effects on IQ Real?

The most stunning claim that emerged from Rosenthal and Jacobson's (1968) study was that teacher expectations have self-fulfilling effects on intelligence. Intelligence change is not just any dependent variable. IQ test scores often are the best predictors of many important life outcomes, including high school and college graduation rates; occupational success, income, and status; and likelihood of becoming an unwed mother or a convicted criminal (Detterman & Thompson, 1997; Neisser et al., 1996). Intelligence

clearly results from an interplay of genetic and nongenetic influences (Neisser et al., 1996). Nonetheless, it has been far easier for research to demonstrate a partial genetic basis for intelligence than to identify the environmental factors that lead to enduring changes in intelligence (Detterman & Thompson, 1997; Neisser et al., 1996).

In this context, the claim that teacher expectations influence IQ was extremely important, controversial, and difficult (for some) to believe. If 40 years of testing various experimental educational programs aimed at reducing disadvantage have struggled mightily to produce enduring increases in IQ scores (Detterman & Thompson, 1997), how likely was it that teacher expectations are endowed with such power? Such a conclusion seemed highly implausible to many researchers in intelligence and standardized testing.

What is the rationale for disputing the effect on IQ? Numerous follow-ups to Pygmalion focusing specifically on the IQ effect have been performed, and both advocates and detractors have addressed this work. Those controversies are reviewed next to identify the light that has been generated by the considerable heat.

The Saga of Wineburg and Raudenbush

In an article titled "The Self-Fulfillment of the Self-Fulfilling Prophecy," Wineburg (1987) provided one of the most sweeping critiques of the IQ effect. First, Wineburg documented how the social and political zeitgeist of the 1960s set the stage for popular acceptance of the study. Next, Wineburg summarized many of the early critiques of the Pygmalion study in a manner that strongly implied they invalidated Pygmalion's conclusions. Nonetheless, even Wineburg (1987, p. 34) recognized the existence of self-fulfilling prophecies: "Within education, the issue had never been whether teachers form expectancies or whether these expectancies affect students." The bone of contention for Wineburg (1987, p. 34) was the effect on IQ: "Obscured and long-forgotten, the heart of the Pygmalion controversy was the bold claim that intelligence was affected by teacher expectations." Wineburg (1987) then reviewed the follow-up studies that focused exclusively on intelligence. That review highlighted the weak-to-nonexistent effect often found on IQ among the follow-ups. Shortly before Wineburg published his article, however, Raudenbush (1984) published a meta-analysis of the effect of experimentally induced teacher expectations on IQ. Wineburg (1987, p. 34) described that meta-analysis as follows: "In a meta-analysis based on 18 studies, Raudenbush (1984) found a small mean effect size in IQ expectation studies ($d = .11$), a finding that either achieved or failed to achieve statistical significance depending on the test employed."

Strictly speaking, there is nothing false here. An effect size of .11 is very small (corresponding to a correlation of about $r = .06$), and Raudenbush did indeed test for statistical significance in several ways, some of which showed that the effect was reliable and some of which did not. But Wineburg's (1987) critique did not address the main point of the Raudenbush's (1984) article. Raudenbush (1984) predicted that the time of year that the study was conducted would moderate expectancy effects. Why? Early in the year, teachers have had little direct experience with their students. In general, all they have is information from their records (previous grades, standardized test scores, etc.) and, perhaps, comments from other teachers. Consequently, they might find new information, such as that provided by a new test of late blooming to be very useful indeed.

In contrast, the later the expectancy induction, the less likely it might be to actually change teachers' expectations. By December, for example, teachers have had extensive contact with their students and have had the opportunity to see for themselves their performance on tests, homework, and in class. Thus, they might be far more likely to discount the importance or validity of a test whose results seemed inconsistent with their direct experience with the student.

Determining this, rather than the overall effect size, was one of the main purposes of Raudenbush's (1984) study. The relationship between time of year of induction and effect size was strongly curvilinear. Effect sizes closely corresponded to the Pygmalion effect of $r = .15$ when the manipulation was conducted within the 1st week of the year, but then rapidly dropped off after that. Expectancy inductions introduced more than 2 weeks into the school year produced no effect. Raudenbush (1984) showed that this curvilinear relationship was highly statistically significant.

The More Recent Exchanges

In 1994, Rosenthal updated the 1978 Rosenthal and Rubin meta-analysis with more recent research but reached essentially the same conclusions as in 1978. In a reply, Snow (1995) emphasized that he agreed that self-fulfilling prophecies were a genuine phenomenon. However, he argued that there was no evidence supporting the hypothesis that teacher expectations influence intelligence. He provided an intriguing re-analysis of the original Pygmalion data, revealing that many of the first and second graders' scores (those among whom the expectancy effect was strongest) were bizarre: Some students had pretest IQ scores near zero, and others had posttest IQ scores over 200. Obviously, however, the children were neither vegetables nor geniuses.

Furthermore, Snow (1995) pointed out that the TOGA was only normed for scores between 60 and 160. If one excluded all scores outside this range, the expectancy effect disappeared. Moreover, there were

five "bloomers" with wild IQ score gains that averaged over 90 points: 17–110, 18–122, 133–202, 111–208, and 113–211. If one simply excluded these five bizarre gains, the difference between the bloomers and the controls evaporated.

Snow (1995) also criticized the conclusion reached in Raudenbush's (1984) meta-analysis. He pointed out that some teacher expectation-IQ studies produced reversals (higher IQ scores or gains in the control group) and argued that the minuscule median effect size (d) of .035 was a better estimate of the effect than the modest but significant mean effect size that Raudenbush reported.

However, at about the same time, Raudenbush (1994) published a re-analysis of the 18 experiments included in his earlier meta-analysis using random effects models, which permit greater generalization than did his earlier method assuming fixed effects. The effect size (d) for the four studies in which there was no prior teacher–student contact was .43, corresponding to a correlation of about $r = .2$ between expectancy manipulation and IQ. (The remaining 14 studies still showed no overall effect.)

Raudenbush (1984, 1994) has concluded that teacher expectations do influence IQ. Snow (Elashoff & Snow, 1971; Snow, 1969, 1995) and Wineburg (1987) have concluded that it does not. Another recent review of this issue (Spitz, 1999) weighed in closer to the Snow/Wineburg position than to that of Raudenbush. It therefore appears that whether teacher expectations have much influence on student intelligence remains controversial and unresolved.

The disagreement and controversy surrounding this issue might convey the impression that no conclusions at all can be reached on the basis of existing data. Such an impression, however, goes too far. The existing research clearly does constrain the range of possible conclusions: Self-fulfilling effects of teacher expectations on student IQ range from nonexistent (if one accepts Snow's and Wineburg's interpretations) to small (if one accepts Rosenthal's and Raudenbush's interpretations). Regardless of which conclusion one believes the data most clearly supports, the following conclusion is certain: The hypothesis that teacher expectations have large and dramatic effects on IQ has been disconfirmed.

This raises at least two questions: (a) Why might expectancy effects be small and unusual? (b) Are small and unusual effects restricted to IQ (i.e., are larger effects found with measures of achievement, rather than intelligence)? These issues are discussed next.

The Accuracy and Power of Teacher Expectations

One possible explanation for why teacher expectation effects may not be particularly powerful is that teachers are generally accurate. Because accurate be-

liefs, by definition, do not create self-fulfilling prophecies (see Merton, 1948), accuracy limits self-fulfilling prophecies. Accuracy is itself a rich and controversial topic and can refer to a wide variety of social phenomena beyond the scope of this review (Funder, 1999; Jussim, 1991; Kenny, 1994). Accuracy, in this article, narrowly refers to teacher expectations predicting but not causing student achievement. Although some researchers consider self-fulfilling prophecy a type of accuracy (Jost & Banaji, 1994; Swann, 1984), it is both conceptually important and empirically possible to distinguish between the specious “accuracy” of expectations that lead to their own fulfillment and the accuracy of expectations that successfully predict some outcome but do not cause that outcome (see also Brophy, 1983; Jussim, 1991; Merton, 1948, for similar discussions).

The answers to these two questions (how accurate are teacher expectations and how powerful are teacher expectation effects) are strongly inversely linked: As accuracy increases, the potential for self-fulfilling prophecies declines; as accuracy decreases, the potential for self-fulfilling prophecies increases. This naturally leads to two empirical questions: How accurate are typical teacher expectations? and How strong are the typical self-fulfilling effects of teacher expectations? As shall be seen, two literatures—one in social psychology and one in educational psychology—have reached diametrically opposed conclusions on this issue. Those literatures are discussed next.

Talking Past Each Other

Social psychologists and educational psychologists have, with a few exceptions, largely talked past each other regarding expectancy effects for most of three decades. There was no direct controversy between the disciplines (e.g., leading educational psychologists never took on leading social psychologists in any sort of exchange of differing viewpoints), but claims regarding the power of self-fulfilling prophecies commonly reached by social psychologists contrasted sharply with claims equally commonly reached by educational psychologists.

Reviews of the teacher expectation literature by educational psychologists typically emphasized the limited power of teacher expectations to influence students (Brophy, 1983; Brophy & Good, 1974; West & Anderson, 1976). In contrast, reviews of the interpersonal expectancy literature (which typically included but were not restricted to teacher expectations) by social psychologists typically emphasized the substantial power and pervasiveness of self-fulfilling prophecies (Fiske & S. Taylor, 1991; Jones, 1986; Miller & Turnbull, 1986; Schultz & Oskamp, 2000; Snyder, 1984).

Indeed, the conflict in basic ideas was even more pronounced, because those same educational psychology reviews typically concluded that teacher expecta-

tions predicted student achievement far more because those expectations were accurate than because they were self-fulfilling. In contrast, not only did social psychological reviews typically focus far more on error, bias, and self-fulfilling prophecy than on accuracy (Fiske & S. Taylor, 1991; Jones, 1990; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980), but many social psychological reviews explicitly disputed the viability and value of even assessing accuracy in social perception (Fiske, 1998; Jones, 1985, 1986, 1990; Stangor, 1995). Consider the following quotes by educational psychologists:

Studies of in-service teachers' expectations for their actual students reveal that most teacher perceptions of students are accurate and based on the best available information, and most of those that are inaccurate are corrected when more dependable information becomes available. (Brophy, 1985, p. 304)

Teacher expectancies exist and they are quite accurate. The effects of teacher expectancies on students are less clear but surely they occur, although not with the frequency or intensity that was suggested by earlier investigators. (Meyer, 1985, p. 361)

Contrast those with the following from social psychologists.

General claims regarding interpersonal expectancies:

Once such an expectation is held about an individual, of course, self-fulfilling prophecy during social interaction should ensure that the hypothesis is behaviorally confirmed. (Skov & Sherman, 1986, p. 116)

Several decades of experimental research in social psychology have been devoted to demonstrating the depths and patterns of inaccuracy in social perception. ... This applies ... to most empirical work in social cognition. ... The thrust of dozens of experiments on the self-fulfilling prophecy and expectancy-confirmation processes, for example, is that erroneous impressions tend to be perpetuated rather than supplanted because of the impressive extent to which people see what they want to see and act as others want them to act. (Jost & Kruglanski, 2002, pp. 172–173)

Claims specifically regarding teacher expectations:

Teachers' expectancies influence students' academic performance to a greater degree than stu-

dents' performance influences teachers' expectancies. (Miller & Turnbull, 1986, p. 236)

Referring to Rosenthal & Jacobson, 1968:

The teachers' expectations had a dramatic impact on the actual performance of the spurters. (Gilbert, 1995, p. 131)

The rejection of accuracy:

The naïveté of this early [accuracy] research was ultimately exposed by Cronbach's elegant critique in 1955. Cronbach showed that accuracy criteria are elusive and that the determinants of rating responses are psychometrically complex. Prior to this pivotal analysis, however, Asch solved the accuracy problem by by-passing it. (Jones, 1985, p. 87)

How we know what isn't so: The fallibility of human reason in everyday life. (Gilovich, 1991, book title)

These quotes are neither unusual nor taken out of context—social psychological reviews and discussions have a long history of emphasizing the inaccuracy of expectancies and their power to create self-fulfilling prophecies (Darley & Fazio, 1980; Fiske & S. Taylor, 1991; Jones, 1986, 1990; Snyder, 1984). This disconnect between the conclusions reached regarding essentially the same phenomenon in two subdisciplines of psychology is both striking and perplexing. The next section, therefore, seeks to identify potential reasons for this disconnect and to evaluate which set of conclusions are most clearly supported by the empirical evidence.

Why the Disconnect Between Social and Educational Psychological Conclusions Regarding Essentially the Same Topic?

Disciplinary boundaries. Although the answer to this question is probably difficult to pin down with certainty, there are several likely suspects. First, there is the common difficulty in bridging barriers between disciplines. Interdisciplinary research is the exception, not the rule. Nonetheless, such barriers are typically somewhat porous (consider, e.g., hybrid domains such as social cognition, cognitive neuroscience, political psychology, etc.), so this does not offer a complete explanation for why this particular disconnect occurred.

Educational psychology's skepticism toward versus social psychology's embracing of self-fulfilling prophecies. Educational psychologists and social psychologists had very different reactions to the initial Rosenthal and Jacobson (1968) Pygmalion study. Social psychological reviews typically accepted the study's conclusions at face value (Fiske & S. Taylor, 1991; Gilbert, 1995; Jones, 1986; Miller & Turnbull, 1986), and they often interpreted the study as a testament to the power of expectations to create social reality along the lines of the simplified version described previously. Furthermore, social psychology often used Rosenthal and Jacobson's (1968) methodology of experimentally testing the effects of experimentally induced false expectations as a model of how to study self-fulfilling prophecies (Darley & Fazio, 1980).

In contrast, the Pygmalion study was greeted with skepticism and some scathing criticisms within educational psychology (Elashoff & Snow, 1971; Jensen, 1969; Snow, 1969; Thorndike, 1968). These criticisms may have dramatically increased educational researchers' perceived accountability. Knowing that the conclusions of subsequent studies and reviews of teacher expectation effects were likely to be subject to high levels of skeptical scrutiny and critical evaluation may have increased educational psychologists' sensitivity not only to basic methodological issues, but also to issues such as alternative explanations, and the size of any self-fulfilling effects obtained. This may help explain the early emergence of accuracy within educational psychology as an alternative explanation for why (except when experimentally induced) teachers' expectations often predict student achievement and the very early focus by educational psychologists on the effect sizes obtained in expectancy studies, which were typically fairly small (see, e.g., Brophy & Good, 1974; West & Anderson, 1976, for early reviews).

Social psychology's rejection of accuracy. Self-fulfilling prophecy research was prominent in the middle of social psychology's three-decade-long (roughly 1955–1985) dismissal of accuracy as a construct non grata. This probably rendered many social psychologists unreceptive to claims suggesting that teacher expectations were typically accurate. This rejection was especially acute among researchers working in the dominant area of social psychology—social cognition—almost to the point of stigmatizing accuracy research (Fiske, 1998; Jones, 1985, 1986, 1990; Stangor, 1995). The 30-year near-banishment of accuracy research within social psychology did not create much firm ground for widespread acceptance of the idea that teachers' expectations might actually predict student achievement primarily because those expectations were accurate.

Differing methodologies of choice lead to differing salience of effect sizes. Differences in preferred research methodologies also probably differentially sensitized workers in the two disciplines to effect size issues. Whereas much educational psychology work focused on naturally occurring teacher expectations and employed structural equation techniques to test their effects (West & Anderson, 1976; Williams, 1976), social psychologists typically relied on experimental laboratory studies (see Darley & Fazio, 1980, for a review). Effect sizes (e.g., standardized regression coefficients) are a common manner in which effects are reported when using structural equation techniques, thereby facilitating the likelihood that such information would be reported in many naturalistic educational psychology studies.

In contrast, throughout the 1970s and 1980s, reports of experimental studies within social psychology rarely reported effect sizes. Instead, the logic of null hypothesis testing of differences between conditions led to starkly dichotomous conclusions: If the difference between expectancy conditions was statistically nonsignificant, no self-fulfilling prophecy occurred; if it was statistically significant, one occurred. Because much of the early social psychological research did yield statistically significant evidence of self-fulfilling prophecies (Darley & Fazio, 1980), in the era prior to concern with effect sizes, conclusions emphasizing their power and pervasiveness probably seemed justified to reviewers of the experimental social psychology research on expectancy effects.

How large are typical teacher expectancy effects? Both meta-analyses and narrative reviews (Brophy, 1983; Jussim, 1991; Raudenbush, 1984; Rosenthal & Rubin, 1978) support the conclusion that, on average, the self-fulfilling effects of teacher expectations are small (averaging about $r = .1$ to $.2$). Because meta-analyses in many areas of research frequently

discover that average effect sizes for particular phenomena are smaller than once believed, one criterion for evaluating the average effect of teacher expectations would involve comparison to effect sizes typically obtained in other meta-analyses. By this criterion, the average teacher expectation effect size falls in the bottom third of effect sizes obtained in 380 meta-analyses (Hemphill, 2003).

Another criterion by which one could evaluate the power of this effect is by translating the effect size into a metric that indicates the proportion of students in a particular class likely to be affected by self-fulfilling prophecies. Twenty years ago, Brophy's (1983) narrative review concluded that, on average, teacher expectations typically have self-fulfilling effects on only 5%–10% of students. This conclusion has held up remarkably well. As shown in Table 1, Rosenthal's (1984) binomial effect size display (BESD) also shows that the typical teacher expectation effect of $.1$ to $.2$ means that self-fulfilling prophecies typically change the achievement of about 5%–10% of all students. Ranges of 5% to 10% can be important, especially for students among that 5%–10%. Obviously, however, it also means that, on average, 90%–95% of the time, students are unaffected by teacher expectations. Clearly, therefore, the wealth of accumulated data on teacher expectations justifies generalizations that emphasize their limited power, rather than generalizations that characterize such effects as dramatic or impressive.

How accurate are teacher expectations? Accuracy has made a comeback within social psychology over the last 10–15 years (Funder, 1999; Ickes, 1997; Jussim, 1991; Kenny, 1994). However, little in this new wave of accuracy research has addressed the issue of teacher expectations. The few exceptions (Jussim, 1989; Jussim & Eccles, 1992; Jussim et al., 1996; Trouilloud, Sarrazin, Martinek, & Guillet, 2002) have consistently shown that teacher expectations predict

Table 1. *The Limited Power of Teacher Expectations to Create Self-Fulfilling Prophecies As Indicated by Rosenthal's Binomial Effect Size Display*

	Low Teacher Expectations (%)	High Teacher Expectations (%)
Teacher expectations have no effect ^a		
Students with above average future achievement	50	50
Students with below average future achievement	50	50
Teacher expectations have an $R = .1$ effect on student achievement ^b		
Students with above average future achievement	45	55
Students with below average future achievement	55	45
Teacher expectations have an $R = .2$ effect on student achievement ^c		
Students with above average future achievement	40	60
Students with below average future achievement	60	40

^aThis table shows that teacher expectations are irrelevant to student achievement. Regardless of whether teacher expectations are high or low, 50% of students end up with above average achievement and 50% end up with below average achievement.

^bA teacher expectation effect of $R = .1$ substantially affects 5% of all students.

^cA teacher expectation effect of $R = .2$ substantially affects 10% of all students.

student achievement primarily because those expectations are accurate.

Within educational psychology, assessing the accuracy of teacher expectations was never viewed as unusually problematic and was accomplished in two main ways. First, the results of studies that simply correlated teacher expectations with student achievement (Brophy & Good, 1974; Hoge & Butcher, 1984; Humphreys & Stubbs, 1977) were compared with the effects of teacher expectations obtained in experimental studies. Such comparisons provided indirect evidence for high accuracy because the correlations were typically much higher (generally in the .4 to .8 range) than were the expectancy effect sizes (standardized regression coefficients typically in the .1 to .2 range—see, e.g., Brophy, 1983; Jussim, 1991, for reviews). The difference between the correlation and the effect size constitutes an indirect way to estimate the accuracy of teacher expectations, because this difference represents predictive accuracy without self-fulfilling influence (Jussim, Eccles, & Madon, 1996). By this metric, about 75% of the overall predictive validity of teacher expectations for standardized test scores reflects accuracy and the remaining 25% reflects self-fulfilling prophecy.

The second way of evaluating the accuracy of teacher expectations was to empirically assess it within a study (rather than compare results across studies). The basic methodology involved (a) assessing teacher expectations (typically early in the school year); (b) assessing student achievement in the year prior to the assessment of teacher expectations; (c) assessing student outcomes at the end of the school year in which teacher expectations were assessed (most typically, standardized test scores, but sometimes, grades, course selections, etc.); and (d) examining the extent to which teacher expectations predicted but did not cause student outcomes.

The logic here is straightforward. The correlation between teacher expectations early in the year and student achievement at the end of the school year represents the overall predictive validity of teacher expectations. That predictive validity can come from only two sources, which are both mutually exclusive and exhaustive: (a) teacher expectations cause student achievement (e.g., through self-fulfilling prophecies); and (b) teacher expectations predict, but do not cause, student achievement. To the extent that both teacher expectations and student achievement are caused by third variables, they will correlate without causing one another. Longitudinal data, of course, precludes the possibility of student end of year achievement causing teacher expectations at some earlier time in the school year.

The standardized path coefficient (whether obtained in regression, LISREL, HLM, or any structural equation technique) linking teacher expectations to

student achievement in the context of a model that controls for plausible sources of accuracy (student prior grades and achievement, demographics, motivation, etc.) represents the best estimate of a naturally occurring self-fulfilling prophecy. It represents the best estimate of the extent to which teacher expectations early in the year predict changes in student achievement by the end of the school year (we know this because prior achievement is controlled). The difference between the overall predictive validity of teacher expectations (the correlation with achievement) and the standardized path coefficient estimating self-fulfilling prophecy equals the extent to which teacher expectations predicted but did not cause, student achievement. Prediction without causation is exactly how we define accuracy (see Alwin & Hauser, 1975, for a discussion of the decomposition of effects in path analysis, and see, e.g., Jussim, 1991, for a detailed example demonstrating how accuracy mathematically and statistically equals the correlation minus the path coefficient linking teacher expectations to students, future achievement).

Of course, naturalistic studies are not experiments. This, however, constitutes a threat not to the accuracy interpretation of such studies, but to the self-fulfilling prophecy interpretation! This is because, no matter how well conducted any naturalistic study is, it is always possible that it has omitted some important third variable. Such a variable, if it exists, would mean that teacher expectations are even less powerful than we have concluded, and that accuracy is even higher than we have concluded. Why? Because it would constitute an alternative explanation for the path coefficient estimating the self-fulfilling effect of teacher expectations on student achievement.

The bottom line, however, has been that studies using this approach yielded essentially the same results as the cross-study comparisons (see reviews by Brophy, 1983; Jussim & Eccles, 1995). Although self-fulfilling prophecies do occur, teacher expectations predict student achievement mainly because those expectations are accurate.

Social psychological testaments to the power of social beliefs to create social reality, and to the inaccuracy of social beliefs, were not based primarily on teacher expectation research. However, such perspectives frequently cite teacher expectation research to justify their conclusions and/or assume that their general conclusions about the power of beliefs to create reality apply to educational contexts (Aronson, 1999; Fiske & S. Taylor, 1991; Jones, 1986, 1990; Jost & Kruglanski, 2002; Myers, 1999; Nisbett & Ross, 1980; Schultz & Oskamp, 2000; Snyder, 1984). To the extent that psychologists make claims based on or applied to teacher expectation effects, the empirical data on teacher expectations should, presumably, both inform and constrain discussions of the inaccuracy and power of beliefs to create reality. In this context, as broad gen-

eralizations, the common social psychological emphasis on inaccuracy and the power of self-fulfilling prophecies, at least with respect to teacher expectations, do not appear well justified either by the original Rosenthal and Jacobson (1968) study or by the subsequent teacher expectation research.

Some social psychological reviews have begun moving closer to the educational psychology perspectives. Many social psychological reviews maintain the field's traditional skepticism regarding accuracy and/or its traditional emphasis on the power and pervasiveness of error, bias, and expectancy effects (Claire & Fiske, 1998; Fiske, 1998; Gilbert, 1995; Jones, 1986, 1990; Jost & Kruglanski, 2002; Oskamp & Schultz, 2000; Stangor, 1995). Nonetheless, at least two recent reviews of interpersonal expectancies, though focusing primarily on sources and effects of expectancies, have presented perspectives at least somewhat closer to those of the educational psychologists. Specifically, with varying degrees of reservation, two recent reviews have acknowledged the evidence that expectancies are often quite accurate and that many conditions exist that limit the occurrence of self-fulfilling prophecies (Olson et al., 1996; Snyder & Stukas, 1998).

Moderators and Evidence of Powerful Self-Fulfilling Prophecies

Although the evidence does not justify broad generalizations emphasizing the power of expectancies, there still may be some conditions in which self-fulfilling prophecies are larger than the typically small effects of .1–.2. The next section, therefore, reviews the evidence on moderators of self-fulfilling prophecies, which includes identifying some conditions under which genuinely powerful self-fulfilling prophecies do occur.

Timing of false expectations. One such situation has already been discussed—Raudenbush's (1984) meta-analysis showing that experimental studies produce a self-fulfilling prophecy effect on IQ of .2, when false teacher expectations are induced within the first 2 weeks of the school year, and of 0 when false teacher expectations are induced thereafter.

Thus, teachers are most susceptible to developing false expectations on the basis of erroneous information provided by experimenters early in the year, prior to having become familiar with the skills and competencies of their students. Such false expectations are likely to have typically small self-fulfilling prophecy effects of about .2. After teachers have gotten to know their students, however (after the first couple of weeks of the year), their expectations are not readily influenced by manifestly false information provided by ex-

perimenters, which, therefore, do not lead to self-fulfilling prophecies.

Age and new situations. Raudenbush (1984) also found that the power of self-fulfilling prophecies varied by grade level. The strongest teacher expectation effects occurred in first, second, and seventh grades. A simple "younger children are more susceptible" hypothesis can account for the Grades 1 and 2 effect, but not for the Grade 7 effect. Another possibility is that people are most susceptible to self-fulfilling prophecies when they enter new situations—and people in general may be more vulnerable to all sorts of social influences in situations with which they are not familiar (see Jussim et al., 1996, for a review). This latter interpretation is also consistent with research showing that some of the most powerful self-fulfilling prophecies ever found have occurred among new military recruits (McNatt, 2000). In the classroom, however, even these relatively more powerful effects between first, second, and seventh graders only averaged to an r of about .2.

Students' perceptions of differential teacher treatment. Self-fulfilling prophecies may occur because teachers behave differently toward high- and low expectancy students. Teachers are typically emotionally warmer and more supportive to their high expectancy students, provide them clearer and more positive feedback, teach them more and more difficult material, and give them more opportunities to demonstrate mastery (see reviews by Brophy, 1983; Jussim, 1986; Rosenthal, 1974; see the meta-analysis by Harris & Rosenthal, 1985).

Based on these patterns, Brattesani, Weinstein, and Marshall (1984) hypothesized that teachers who engage in more differential treatment of their students would also produce larger self-fulfilling prophecies. To identify such teachers, Brattesani et al. (1984) asked students to indicate how much their teachers treated different students differently. Brattesani et al. (1984) then split the teachers into two groups: Those who students identified as engaging in much differential treatment; and those who students identified as engaging in little differential treatment. Analyses then assessed the extent to which teacher expectations predicted student self-expectations and their future achievement, after controlling for achievement the prior year.

As anticipated, teacher expectations most strongly predicted student expectations and achievement in classes where the students perceived the greatest differential treatment. The effect sizes for teacher expectations predicting student expectations and achievement ranged from about 0 to .1 among the low differential treatment classes, and from about .3 to .4 among the high differential treatment classrooms.

Tracking by ability level. School tracking refers to the policy of segregating students into different classes according to their ability. For example, smart students may be assigned to one class, average students to another, and slow students to a third. Because tracking represents institutional justification for believing that some students are more able than others, some researchers have suggested that tracking may lead to the type of rigid teacher expectations most likely to create self-fulfilling prophecies (Oakes, 1986).

The one study to empirically investigate this hypothesis, however, failed to support it (Smith et al., 1998). Self-fulfilling prophecies among students grouped by ability between classes were no more powerful than those among students in heterogeneous classes—and both fell within the typically small range of 0–.2 (the results varied slightly by predictor and outcome, but not by between class groups).

There was, however, some evidence that within-class grouping moderated self-fulfilling effects of teacher expectations (within-class grouping refers to the practice of dividing students into two or more ability groups within a class). Although effects were near zero among students who were either not grouped at all, or who were in high groups, such effects were about .2 among those in low ability within class groups. Ability differences and group labels may be more salient to teachers who use within-class grouping. This may increase differential treatment, leading to greater self-fulfilling prophecies. These results mean, however, that, although there was evidence of moderation, there was also no evidence of atypically large self-fulfilling prophecies.

Student stigmatization. Students who belong to a stigmatized group may be particularly vulnerable to self-fulfilling prophecies. Membership in stigmatized groups has special importance among the possible moderators, because of its obvious relevance to the perpetuation of social inequalities. Yet, despite its importance, the role of stigmatized status in moderating classroom self-fulfilling prophecies has only been addressed by two studies. One examined whether self-fulfilling prophecies were stronger among students with prior histories of high or low achievement (Madon, Jussim, & Eccles, 1997). Consistent with the prediction that the socially stigmatized are more strongly affected by self-fulfilling prophecies, Madon et al. (1997) found a self-fulfilling prophecy effect size among low achievers of .26, whereas the self-fulfilling prophecy effect size among high achievers was only .08.

Another study examined whether teacher expectations produced stronger self-fulfilling prophecies among students from stigmatized demographic groups (Jussim et al., 1996). Even though they studied math classes (and beliefs about girls' alleged inferiority at math are widespread), they tested for but found no evi-

dence that student sex moderated self-fulfilling prophecies. They did, however, find evidence of moderation by social class and race–ethnicity.

Although there was no consistent evidence of self-fulfilling prophecies among students from higher socioeconomic backgrounds, teacher expectations did produce self-fulfilling prophecy effects of .2–.3 among students from lower socioeconomic backgrounds. Although even these effects were not particularly large, there was evidence that teacher expectations for low achieving students from lower social class backgrounds produced a self-fulfilling prophecy effect size of about .6. Furthermore, teacher expectations for African American students produced self-fulfilling prophecy effect sizes of .4 to .6.

The self-fulfilling prophecies among lower achieving students from lower socioeconomic backgrounds and among African American students are large by any standard. These results are broadly consistent both with two traditional social psychological emphases regarding expectancy effects: their potential power and their potential role in social problems.

Because it might seem obvious that such patterns reflect effects of erroneous stereotypes, two follow-up studies examined the accuracy of teacher expectations for African American students and for students from lower social class backgrounds (Jussim et al., 1996; Madon et al., 1998). Instead of finding inaccuracies, however, they found that teachers perceived differences between different groups that closely corresponded to those groups' actual differences in prior grades and achievement tests. Although such findings appear to conflict with narrative reviews emphasizing the inaccuracy and biasing effects of stereotypes (American Psychological Association, 1991; Aronson, 1999; Jones, 1986, 1990), they are consistent with meta-analyses showing that biasing effects of stereotypes on person perception judgments tend to be quite small, typically averaging to an *r* of about .1 (Davison & Burke, 2000; Kunda & Thagard, 1996; Mazella & Feingold, 1994; Sweeney & Haney, 1992; Swim, Borgida, Maruyama, & Myers, 1989).

Conclusions about moderators. Considering the importance of understanding the extent to which self-fulfilling prophecies perpetuate social inequalities, surprisingly little research has been conducted regarding students' social backgrounds as moderators of self-fulfilling prophecies. Further research regarding these and other potential moderators of the power of teacher expectancies is clearly needed. This is especially true because theoretical and methodological limitations to studies conducted to date may qualify the validity or generalizability of their findings. Except for Raudenbush's (1984) meta-analysis, only a single study has examined each of the moderating conditions reviewed here. Each of the individual studies focused

on only a limited number of grade levels and subject matters (and often only one), and none were conducted among nationally representative samples, so that generalizability to students from differing demographic backgrounds living in different geographic areas and in different educational and societal contexts is unclear. Therefore, whether other researchers can replicate any of the moderating conditions identified by the individual studies is an open empirical question.

The evidence of moderation that does exist, however, provides a mixed picture regarding the possibility that the self-fulfilling effects of teachers' expectations contribute to social problems. Arguing against such a possibility is the evidence that teacher expectations are generally accurate, and that self-fulfilling prophecies are generally small. Arguing for such a possibility is the evidence that substantial self-fulfilling prophecies occur more frequently among students from stigmatized backgrounds. The social problems interpretation of self-fulfilling prophecies would be further bolstered if these larger effects among stigmatized were consistently more harmful than helpful. The next section, therefore, reviews the evidence regarding the relative power of positive and negative teacher expectations.

Do Negative Teacher Expectations Harm More Than Positive Teacher Expectations Help?

Even typically small teacher expectations effects could still be major contributors to social problems and injustices, if such effects were nonlinear. Perhaps the effects of .1 to .2 obtained in most naturalistic studies and meta-analyses, nearly all of which assessed linear relations between teacher expectations and student achievement, reflect effects of about .3 to .4 for negative teacher expectations and effects of near zero for positive teacher expectations. If teacher expectations consistently and substantially harm the achievement of low expectancy students, but rarely lift up the achievement of high expectancy students, they would be a potent and harmful phenomenon.

It is at least hypothetically possible, however, that the self-fulfilling effects of teachers' expectations might be more helpful than harmful. If so, they might alleviate social problems more than they contribute to such problems. Perhaps positive teacher expectations create larger self-fulfilling prophecies than do negative teacher expectations. Perhaps positive teacher expectations have effects of .3–.4 and negative teacher expectations effects near zero. An overall linear effect of .1–.2 could also obtain if this were true.

Whether self-fulfilling prophecies are primarily benevolent and usually enhance students' achievement or primarily harmful and usually undermine students' achievement is not a controversy. However, this is an

odd case where perhaps there should be some controversy, because (a) many reviews have explicitly argued or implied that self-fulfilling prophecies are most likely to have negative effects (Darley & Fazio, 1980; Devine, 1995; Fiske & S. Taylor, 1991; Gilbert, 1995; Jones, 1986, 1990; Weinstein & McKown, 1998) and (b) There have been only three published teacher expectations studies on this issue, which have yielded a decidedly mixed picture. Those studies are discussed next.

Do “High Bias” Teachers Produce More Negative Self-Fulfilling Prophecies in Gym?

Babad, Inbar, and Rosenthal (1982) examined the power of negative and positive self-fulfilling prophecies among 26 teachers and 202 students in gym classes who had either low bias or high bias teachers (bias referred to degree of cognitive rigidity or dogmatism among teachers). This study reached the conclusion that negative self-fulfilling prophecies were more powerful than positive ones, at least among high bias teachers. Whether the study actually provided the evidence necessary to justify this claim, however, is subject to some doubt.

Babad et al. (1982) found no differences in athletic accomplishments between high- and low expectancy students' performance among low bias teachers—that is, no self-fulfilling prophecy. In contrast, they did find that the high expectancy students performed more highly than did the low expectancy students among high bias teachers (demonstrating occurrence of a self-fulfilling prophecy). However, a difference between high- and low expectancy students is insufficient to determine whether self-fulfilling prophecies primarily helped or harmed students. This difference could occur if (a) high expectations helped students and low expectations had no effect, (b) low expectations harmed students and high expectations had no effect, or (c) high expectations helped students and low expectations harmed students.

Because there was no evidence that low bias teachers induced self-fulfilling prophecies, students' performance among low bias teachers could be used as a sort of control group for determining whether self-fulfilling prophecies primarily helped or hurt students with high bias teachers. Among students with high bias teachers, if negative self-fulfilling prophecies were more powerful than positive self-fulfilling prophecies, then (a) low expectancy students with high bias teachers should have consistently performed worse than low expectancy students with no bias teachers and (b) there should be little difference between the performance of high expectancy students with high or no bias teachers.

This was the case for only one of three main dependent variables. The three student performance measures were (a) distance jump, (b) sit-ups (for girls) and push-ups (for boys), and (c) running speed. For the distance jump, negative self-fulfilling prophecies were more powerful than positive ones. Lows with high bias teachers jumped significantly less far than did lows with no bias teachers; highs with high bias teachers jumped the same distance as highs with low bias teachers.

For sit-ups/push-ups, positive self-fulfilling prophecies were more powerful than negative ones. Low expectancy students of high bias teachers performed 3.8 fewer sit-ups/pushups than did low expectancy students of no bias teachers; high expectancy students of high bias teachers performed 4.7 more sit-ups/pushups than did high expectancy students of low bias teachers (see their Table 5, p. 469).

The speed measure also provided no evidence of negative expectancy effects exceeding positive ones. The performance of low expectancy students with no bias teachers and high bias teachers were similar, indicating that negative self-fulfilling prophecies did not occur. High expectancy students with high bias teachers actually performed worse than did high expectancy students with no bias teachers, which may be an interesting effect of teacher bias, but does not represent a self-fulfilling prophecy.

Overall, Babad et al.'s (1982) results provided a mixed picture. They found evidence of both negative and positive self-fulfilling prophecies. Their research did not provide evidence that negative self-fulfilling prophecies were consistently stronger than positive self-fulfilling prophecies.

Under-Estimating Versus Over-Estimating IQ

Sutherland and Goldschmid (1974) assessed six first- and second-grade teachers' expectations 2 months into the school year. Ninety-three students were divided into five teacher expectation groups (ranging from "poor" to "superior"). The students were administered intelligence tests at each of two time points: 2 months and 7 months into the school year.

Sutherland and Goldschmid (1974) first focused on students with below average IQ scores, who were divided into two groups: (a) those whom teachers believed had average intelligence (erroneously high expectation) and (b) those whom teachers believed had below average intelligence (accurately low expectation). The self-fulfilling prophecy prediction is that students in the first group (low student IQ/inaccurately high teacher expectation) would show greater increases in IQ over the year than students in the second group (low student IQ/accurately low teacher expecta-

tion). The pattern of increases confirmed the prediction for both IQ tests, but the difference was not statistically significant (effect sizes of .1 to .2).

Next, Sutherland and Goldschmid (1974) divided students with above-average IQ test scores into two groups: (a) those whom teachers believed had above-average intelligence (accurately high expectation) and (b) those whom teachers believed had average intelligence (inaccurately low expectations). The self-fulfilling prophecy prediction here was that students in the second group (high IQ/inaccurately low teacher expectations) would show less IQ gain than would students in the first group (high IQ/accurately high expectations). This prediction was confirmed for both measures; in addition, these differences were statistically significant and strong (r 's of .45 to .55). These results suggest that negative expectations undermined the IQ scores of high IQ students, whereas positive expectations did not significantly raise the IQ scores of low IQ students.

This study, however, suffered from several methodological weaknesses. First, negative expectations underestimated high IQ students more than positive expectations overestimated low IQ students. Positive expectations consisted of rating as "average" students with IQ scores of 80–95. Negative expectations consisted of rating as "average" students with IQ scores of 120–135. An average IQ score is 100. Thus, an "average" rating probably underestimates a student with a score of 120–135 more than it overestimates a student with a score of 80–95.

The greater power of negative versus positive self-fulfilling prophecies, therefore, may have reflected the greater inaccuracy of negative expectations as operationalized among their particular sample, rather than any generally greater power of negative expectations. More inaccurate expectations have greater potential to be self-fulfilling. Therefore, even if the self-fulfilling effects of teacher expectations in Sutherland and Goldschmid's (1974) data were completely linear, operationalizing teacher inaccuracies so that low expectations would be more inaccurate than high ones would lead to finding that negative self-fulfilling prophecies exceed positive ones.

In addition, the study did not examine the effects of inaccurately low expectations on low IQ students or of inaccurately high expectations on high IQ students. A teacher could believe that some slightly below-average students are even less competent than indicated by their IQ score; or that some high IQ students are even more competent than indicated by their IQ test score. Such effects, however, were not assessed.

Therefore, this study's conclusions can best be summarized as follows: Highly inaccurate low expectations undermine high IQ students' future IQ test scores moreso than moderately inaccurate high expectations

enhance low IQ students' future test scores. Such a specific and narrow conclusion does not appear to provide a firm empirical foundation for broad conclusions regarding the relative power of positive and negative teacher expectations.

Under- Versus Overestimating Achievement in Math Classes

The third study to address the relative power of positive versus negative teacher expectations (Madon et al., 1997) (a) focused on 98 teachers and 1,539 students in sixth grade math classes, (b) explicitly compared inaccurately low expectations to equally inaccurate high expectations, and (c) performed this comparison both overall and separately for high and low achieving students.

To compare the power of positive versus negative expectancy effects, Madon et al. (1997) assessed the degree and direction of teacher inaccuracies relative to each student through regression analyses. Residuals were obtained from an analysis that used students' prior achievement and motivation to predict teacher perceptions early in the year. A positive residual meant that the teacher rated that student more highly than other students with similar prior motivation and achievement ("teacher overestimates"); a negative residual meant that the teacher rated that student less highly than other students with similar prior motivation and achievement ("teacher underestimates").

Polynomial regression (Judd & McClelland, 1989) was then used to test whether teacher over- or underestimates more strongly predicted changes in students' math achievement. The slope of the relationship of teacher expectations to student achievement was about .3 among the most highly overestimated students and about .1 among the underestimated students. This pattern indicated greater power of positive than of negative self-fulfilling prophecies.

In addition, Madon et al. (1997) examined this pattern separately for students with prior records of high or low achievement. For high achievers, teacher underestimates had almost no self-fulfilling prophecy effect and teacher overestimates produced self-fulfilling prophecy effects of about .2. For low achievers, teacher underestimates produced self-fulfilling prophecy effects of about .1-.2 and teacher overestimates produced self-fulfilling prophecy effects of about .4. In sum, Madon et al. found that positive expectancies were more powerful than negative expectancies, and this was especially true for low-achieving students.

Conclusion: Are Positive or Negative Teacher Expectations More Powerful?

It appears that Babad et al. (1982) found no clear and consistent pattern and that Sutherland and

Goldschmid's (1974) study was biased in the direction of finding stronger negative expectancy effects. Madon et al. (1997) provided an unbiased test of the power of positive versus negative self-fulfilling prophecies and examined this issue among a much larger sample than was included in the prior studies. They found that positive expectancy effects were generally more powerful than negative ones, and this pattern disproportionately benefited low expectancy students. Whether the evidence from these three studies tilts in favor of the power of positive expectations or of negative expectations, therefore, is currently a matter of individual scientific judgment.

It is, of course, possible that some conditions facilitate the occurrence of positive self-fulfilling prophecies and others facilitate the occurrence of negative self-fulfilling prophecies. If so, the sparse evidence on this issue does not yet shed light on just what those conditions might be. A potentially rich area for future research, therefore, involves identifying conditions that facilitate positive self-fulfilling prophecies and those that facilitate negative self-fulfilling prophecies, both in general, and among specific subgroups of students (such as the stigmatized).

Disproportionate effects of negative self-fulfilling prophecies, however, are not strictly necessary to maintain the claim that teacher expectations contribute to social problems and inequalities. Even if teachers' expectations never harmed students at all, positive expectancy effects, alone, could create ever-increasing differences between high and low expectancy students, if the same students were the beneficiaries of positive expectancy effects year-in and year-out. Therefore, both theory and evidence regarding whether self-fulfilling prophecies produced by teacher expectations accumulate are examined next.

Do Self-Fulfilling Prophecies Accumulate?

The Logic of Accumulating Self-Fulfilling Prophecies

Many reviews and perspectives have suggested that empirical studies underestimate self-fulfilling prophecies, because expectancy effects may accumulate over time and/or over multiple perceivers (Claire & Fiske, 1998; Jones, 1990; Snyder, 1984; Weinstein & McKown, 1998). The logic of accumulation is straightforward: (a) Small effects are typically obtained in both short-term (1 hr) laboratory studies of self-fulfilling prophecies and teacher expectation studies conducted over a school year; (b) although small in such contexts, many targets may be subjected to the same or similar erroneous expectations over and over again.

For example, students from privileged socio-demographic backgrounds may consistently benefit from high teacher expectations, whereas those from culturally stigmatized backgrounds may be consistently undermined by low teacher expectations. Social stereotypes are often presented as reason to predict that targets from stigmatized groups will be subjected to repeated self-fulfilling prophecies from multiple perceivers and over long periods of time (Claire & Fiske, 1998; Darley & Fazio, 1980; Deaux & Major, 1987; Jones, 1986, 1990; Jost & Banaji, 1994; Snyder, 1984; M. Taylor, 1992). According to this analysis, the cumulative effects of expectancies on a target across contexts or over time are likely to be higher than expectancies demonstrated in any single study.

The logic of accumulation may appear sufficiently obvious as to not require empirical test. Because it is so compelling, it has probably contributed to perspectives emphasizing the power of self-fulfilling prophecies. But before foreclosing on the need to obtain evidence, it might be worthwhile to consider the social psychological processes that could work against accumulation.

Potential Limitations to Accumulation

Because myriad social and psychological processes might work against accumulation, only a few contenders will be briefly mentioned. Within social psychology, perhaps the most obvious is self-verification (Swann, 1987), which refers to the idea that people are highly motivated to see themselves in a manner consistent with their own long-standing and deep-seated self-views, and to convince others to view them much as they view themselves. The self-verification motive may render many people resistant to confirming others' inaccurate expectations.

The only study to address this process in an educational context showed that students self-verify (convince teachers to view them much as they see themselves) to about the same extent that teachers' expectations influence student self-concepts—and both effects were quite small—around .1 (Madon et al., 2001). Furthermore, in a laboratory study conducted over three sessions, Swann and Ely (1984) found that, although self-fulfilling prophecies occurred, targets were more likely to convince perceivers to change their expectations than targets were to fulfill perceivers' expectations. Overall, rather than accumulating, the self-fulfilling effects of perceivers' expectations declined over the three sessions. Thus, self-verification constitutes one potential obstacle to the relentless fulfillment of others' expectations.

A second potential limitation is accuracy. As people get to know one another, the potential to maintain highly erroneous views of one another may decline (al-

though it probably does not decline to zero; see, e.g., Kenny, 1994). Similarly, popular cultural mythology notwithstanding, rather than rigidly applying stereotypes to every individual who is a member of the stereotyped group, people are typically highly sensitive to individual differences, when those individual differences are experimentally manipulated or readily available in naturally occurring situations such as classrooms (Jussim et al., 1996; Kunda & Thagard, 1996). At least two meta-analyses have shown that stereotype effects on judgments of individuals become progressively smaller the more information perceivers have regarding those individuals (Davison & Burke, 2000; Eagly, Makhijani, Ashmore, & Longo, 1991). Stereotype disconfirming behavior is far more likely to be noticed and to influence perceptions and judgments than it is to be ignored or dismissed (Jussim et al., 1996; Kunda & Thagard, 1996). This, too, will typically increase the accuracy of expectations for individuals. If accuracy increases over time, it will limit and reduce the potential for self-fulfilling prophecy.

A third potential limitation is regression to the mean. If a perceiver holds an unusually high or low expectation for a target, even if that expectation is self-fulfilling, the target's behavior may drift back to its pre-self-fulfilling prophecy levels. Regression to targets' prior levels of achievement may create a tendency for self-fulfilling prophecies to dissipate, rather than accumulate.

However, the bottom line is data, not argument. To what extent do the self-fulfilling effects of teacher expectations accumulate? Addressing this question requires understanding two potentially very different aspects or types of accumulation. The first involves accumulation of self-fulfilling prophecies resulting from multiple perceivers within the same time frame (e.g., multiple teachers during the school year). The second involves accumulation of self-fulfilling prophecies over time (e.g., the same teacher over multiple semesters or multiple teachers over multiple years). Both of these issues are discussed next.

Concurrent Accumulation of Expectancy Effects

The accumulation of self-fulfilling prophecies from multiple perceivers' expectations within a single time period. Within a single time frame (e.g., one school year), the effects on targets of multiple perceivers' expectations may accumulate (this is just as true outside of school contexts as inside such contexts, so the generic terms *perceiver* and *target* are used here). To distinguish such effects from the accumulation of expectancy effects over time (e.g., multiple school years), these effects are referred to here as "concurrent accumulation effects." The notion of concurrent accumulation is implicit in most perspectives that

emphasize the potentially self-fulfilling nature of social stereotypes (Claire & Fiske, 1998; Deaux & Major, 1987; Hamilton et al., 1990; Jones, 1990; Snyder, 1984). Because stereotypes are often shared, perceiver after perceiver will presumably heap self-fulfilling prophecy after self-fulfilling prophecy upon stereotyped targets.

This perspective has been explicitly articulated by Claire and Fiske (1998):

To understand the significance of the pressure on targets, one must take the perspective of a target across time and interactions. ... But in constraining possible social influence to short-term one-on-one interactions, the methodology itself reinforces an individualist view of behavior by ignoring the repetitiveness of a target's experience over time and across situations, and the cumulative effect of these interactions. (p. 208)

And later, on p. 211, "Thus, stereotypes are not only widely shared, but some are also pervasively applied in interactions with targets."

The upshot of this analysis is clear: Because all previous research has focused on the potentially self-fulfilling effects of only one perceiver on each target, if multiple perceivers influence targets in daily life, people would be more heavily influenced by self-fulfilling prophecies than is implied by existing research.

A faulty implication. This perspective draws a faulty implication from two sound premises. The two sound premises are as follows: (a) Existing research on self-fulfilling prophecies has focused on dyadic interactions of limited duration; (b) to the extent that targets interact with many perceivers who share expectations, the cumulative effect of self-fulfilling prophecies exceeds such an effect obtained in any given dyadic interaction, which leads to the seemingly self-evident but nonetheless false conclusion that (c) studies focusing on dyadic interactions of limited duration underestimate the extent to which targets are affected by self-fulfilling prophecies, because such studies exclude all perceivers other than those involved in the study.

How can point 3 be false, when points 1 and 2 are true? Point 3 describes the actual state of affairs precisely backward: Studies focusing on dyadic interactions do not underestimate expectancy effects from multiple perceivers; instead, such studies overestimate the effects of any single perceiver's expectations precisely because these studies (unintentionally) incorporate the effects of all other perceivers who hold similar self-fulfilling expectations! This is a self-fulfilling prophecy variant on the well-known "omitted variable problem" in regression (excluded perceivers' expectations are omitted variables). Because the logic of this

analysis is counterintuitive and nonobvious, it is next explicated at some length.

Focus on naturalistic, not experimental, studies.

This analysis focuses exclusively on naturalistic studies for several reasons. The logic of accumulation across multiple perceivers requires that perceivers develop their inaccurate expectations regarding a target spontaneously, and not through experimental intervention. If perceivers rarely spontaneously develop similarly inaccurate expectations, there is not much potential for accumulation in daily life. Contrasting positive and negative expectations, if self-fulfilling, will negate one another rather than accumulate.

Furthermore, Claire and Fiske's (1998) analysis is most fitting for experimental studies of self-fulfilling prophecies. Such studies are typically conducted in very narrow contexts—typically a dyadic interaction that takes place over an hour or less. Even the rare exception, such as the long-term field experiment of Rosenthal and Jacobson (1968), could not address concurrent accumulation effects, because false expectations were experimentally manipulated. Assuming the random assignment to condition was successful, there is no reason to think that other teachers typically held the same expectations for the "late bloomers" as did those in whom Rosenthal and Jacobson (1968) instilled false expectations. Exactly as Claire and Fiske (1998) argued, therefore, such studies do ignore "the repetitiveness of the target's experience over time."

In contrast, the typical naturalistic study often provides an appropriate context for studying the accumulation of concurrent self-fulfilling prophecies. Most naturalistic studies of teacher expectations are conducted over at least 1 school year, thereby allowing for the possibility that multiple teachers will develop similar expectations for students, and at least raising the possibility of concurrent expectancy effects.

Other perceivers' expectations as omitted variables: Why dyadic studies overstate dyadic self-fulfilling prophecy effects and precisely estimate multiple perceiver effects.

Dyadic studies of naturally occurring self-fulfilling prophecies implicitly assess the self-fulfilling effects of all perceivers holding expectations similar to those of the perceiver included in that study, including the potentially infinite number of perceivers not included in the study. The idea that a study can assess effects of perceivers' expectations not included in that study, at first glance, might appear inconceivable and perhaps even nonsensical. It is true nonetheless and here is why.

To illustrate the basic ideas, this section uses a very simple three-variable model and requires a very basic understanding of the decomposition of effects in path analysis (Alwin & Hauser, 1975). The principles, however, apply identically in more complex contexts in-

volving more complex models. In Figure 1, Model 1 two X variables ($X1$ and $X2$) predict Y . $r1$ is the correlation between $X1$ and $X2$; Path A is the effect of $X1$ on Y ; Path B is the effect of $X2$ on Y . All effects here are standardized regression coefficients. In this model, the correlation of $X1$ with Y equals Path A + ($r1$ *Path B). If one excluded $X2$, therefore, and merely estimated the effect of $X1$ on Y , it would exactly equal the correlation of $X1$ with Y . Such an “effect” would be too high by precisely $r1$ Path B. In this example, $X2$ is an “omitted variable” whose omission artificially increases the observed causal effect of $X1$ on Y .

Model 2 presents a concrete example involving teacher expectations and student achievement. This model assumes that two teachers each have an expectancy effect of .2 and that the two teachers hold highly similar ($r = .7$) expectations. In this model, the zero order correlation between One Teacher’s Expectations and Target Students’ Achievement equals $.2 + (.7*.2) = .34$. If one failed to include the Second Teacher’s Expectations in the model, Path A becomes the zero order correlation between One Teacher’s Expectations and Target Students’ Achievement—it equals .34.

This analysis shows ways in which those speculating that expectancy effects are more powerful than they seem are both correct and incorrect. Such speculations are correct in stating that, if two perceivers hold similar self-fulfilling expectations, there will be more of an expectancy effect than if there is only one perceiver holding self-fulfilling expectations. In the concrete hypothetical example, .34 (effect of both teachers’ expectations combined) exceeds .2 (the effect of either teacher’s expectation, by itself).

Nonetheless, the conclusion that existing empirical research underestimates expectancy effects because it

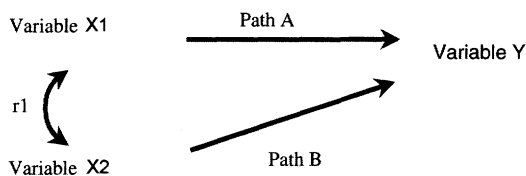
fails to account for multiple perceivers is incorrect. The Figure 1 models demonstrate that studies focusing on dyadic interactions do not underestimate cumulative expectancy effects from multiple perceivers. When Model 2 is the true model, an imperfect model that only included one teacher precisely captures the total accumulation across the two teachers. The flaw in the imperfect, one-teacher model is not that it underestimates accumulation. Instead, its flaw is that it overestimates accumulation. Excluded perceivers are omitted variables that inflate the expectancy effect obtained for the included perceiver.

Omitted variables positively correlated with two included variables will always artificially increase the size of the assessed causal relation between the included variables. The effects of expectations similar to those of the perceiver in the study, as held by the potentially infinite number of perceivers excluded from any particular study, are omitted variables. In other words, concurrent accumulation effects are already (implicitly) assessed in dyadic studies of naturally occurring social interactions, such as between teachers and students.

The empirical evidence on the power of concurrent accumulation effects. No published studies have explicitly assessed the cumulative self-fulfilling effects of multiple teachers’ expectations. Nonetheless, regarding identifying the extent and power of concurrent accumulation effects, the models in Figure 1 show that such studies are not necessary. We already have clear evidence about the extent of such effects. Concurrent accumulation effects exactly equal the effects of self-fulfilling prophecies as assessed in naturalistic studies of dyadic interactions. Such studies overestimate dyadic self-fulfilling prophecies precisely to the extent that concurrent accumulation occurs. This may be a “flaw” in those studies, but regarding understanding the likely extent of concurrent accumulation effects, this flaw is a boon. It means that existing naturalistic research fully identifies the likely extent of concurrent accumulation effects.

This analysis means that the research reviewed earlier on the limited power of naturally occurring dyadic self-fulfilling prophecies effects provides a good estimate of the total effect of all teachers’ (i.e., including those not in the study) expectations that are similar to those of the teachers actually included in the study. Although this review of conceptual issues and empirical evidence cannot conclusively demonstrate the existence of concurrent accumulation effects, it does conclusively demonstrate that, if they occur, they are fully captured by the .1 to .2 self-fulfilling prophecy effect sizes typically found in naturalistic studies of teacher expectations.

Model 1: General Model



Model 2: Two Teachers' Hold Similar Expectations

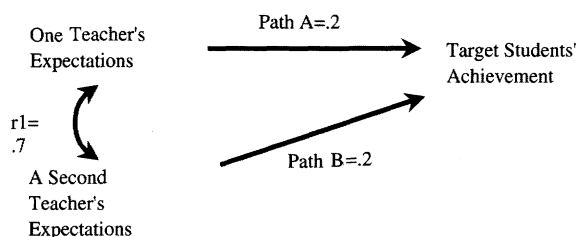


Figure 1. Models of concurrent accumulation of expectancy effects.

That concurrent accumulation effects in the classroom are not particularly large, at least over a school year, naturally leads to the next question: Do such effects accumulate over time?

Accumulation Over Time

The accumulation-over-time hypothesis is that a self-fulfilling prophecy process triggered by a perceiver's expectations at one time continues so that targets conform more and more to the perceiver's original expectations. A perceiver's initial false belief more strongly influences targets over time. Thus, the impact of self-fulfilling prophecies may transcend the original context of the interaction and profoundly influence targets (Claire & Fiske, 1998; Snyder, 1984).

The logic of accumulation over time is, at first glance, as compelling as that of concurrent accumulation: (a) Self-fulfilling prophecies clearly occur in limited contexts, such as the lab or school year, and (b) small effects may accumulate over multiple school years, so that (c) initially small differences between high and low expectancy students become large.

For example, consider two students both starting sixth grade with IQs of 100. Suppose that the sixth grade teacher believes that one of these students is bright and the other is not. Also assume that teachers' expectations have an effect of .2 on the two students' achievement (an effect of .2 is equivalent to one fifth of a standard deviation and the standard deviation of many IQ tests is about 15). Thus, by the end of sixth grade, the "bright" student's IQ will be 103 and the "dull" student's IQ will be 97. If this small effect accumulates over time, then by the end of high school, the bright student will have an IQ of 115, and the dull student will have an IQ of 85. Each year from 6th through 12th grade, the gap between the low and high expectancy students widens by 3 points. Thus, small expectancy effects have the potential to become much more powerful via accumulation.

Again, however, such an analysis is most compelling only in the absence of a comparable analysis of factors likely to limit accumulation over time. Self-verification, accuracy, and regression all may limit accumulation over time. Furthermore, different teachers in different school years may not hold equally inaccurate expectations for most students.

For example, consider a ninth-grade student who would otherwise receive a B in math if he or she was not the target of an erroneously high teacher expectation. Through self-fulfilling prophecies, such a student instead receives a final grade of B+ in math. For this expectancy effect to accumulate, the 10th-grade math teacher would have to have another erroneously high expectation, which is also self-fulfilling, such that the student ends up with an A. If the 10th-grade teacher expected a B+, regardless of whether one considers this

accurate or self-fulfilling, and the student receives a B+, then there is no accumulation. There is merely a self-fulfilling prophecy effect that occurred in ninth grade that was sustained, not increased, in 10th grade.

Again, however, the bottom line is the data. What has research shown regarding the accumulation of self-fulfilling prophecies in the classroom? The four studies that have addressed this issue are discussed next.

Rosenthal and Jacobson (1968)

The classic Pygmalion study showed that self-fulfilling prophecies did not accumulate. Differences between the falsely labeled late bloomers and the accurately labeled controls, at the end of Year 1, were about 4 IQ points; such differences were less than 3 IQ points at the end of Year 2. Thus, self-fulfilling prophecies dissipated (although whether they would have dissipated to zero is a question that cannot be answered by their data, because they only followed students for 2 years). Rosenthal and Jacobson (1968) did not, however, report whether the bloomer/control IQ differences at the end of Year 1 were statistically significantly larger than those at the end of Year 2. Thus, all that can be claimed is that their pattern supported dissipation, not that they had significant evidence of dissipation.

Rist (1970)

Rist (1970) performed an observational study of inner-city children from kindergarten through second grade. He found that, by the eighth day of class, a kindergarten teacher had divided her class into three groups—a supposedly above-average, average, and below-average group. Each group sat at its own table (Tables 1, 2, and 3, respectively). Rist (1970) used table assignment as a criterion for identifying the self-fulfilling effects of teachers' expectations on students. Inasmuch as table assignment constitutes evidence of a teacher decision, rather than direct evidence of student achievement, it is not clear that it is a dependent variable appropriate for assessing self-fulfilling prophecy (which requires a bona fide change in student performance or achievement). Even if one accepts table assignment as an appropriate criterion, however, his results provided some evidence of dissipation and no evidence of accumulation.

In first grade, the teacher placed students from Table 1 in kindergarten at Table A (high group) and all of the students from Tables 2 and 3 at Table B (middle group). Only one of the students from the kindergarten class was placed at the lowest table, Table C, which was comprised mostly of students repeating the grade. At this first transition, differences among students based on table assignment had declined. Although students from the high-ability table remained at a

high-ability table, the students from the middle- and low-ability tables in kindergarten were combined into one middle-ability table in first grade.

In second grade, students from Table A were assigned to the “Tigers” (high group) and students from Table B and C were assigned to the “Cardinals” (middle group). None of the students from the first-grade class were assigned to the “Clowns” (low group). All of the students (who were left—some had moved out of district) who had been in the lowest group in kindergarten were now at least in the middle group. In addition, that year, two students from the Tigers were moved down to the Cardinals and two students from the Cardinals were moved up to the Tigers. Although the groups created by the kindergarten teacher did remain somewhat intact from year to year, by the end of second grade, the initial table-assignment differences between students had decreased.

Rist (1970) interpreted these patterns as indicating that teacher expectations helped to create a caste system based on social class. Nonetheless, the actual results, based on table assignment, provided evidence of both stability and dissipation of self-fulfilling prophecies from kindergarten through second grade. There was no evidence of accumulation.

West and Anderson, (1976)

West and Anderson (1976) analyzed data from 3,000 male students in their freshman, sophomore, and senior years of high school that included information on both teachers’ expectations and student achievement. Overall, the results suggested dissipation. The path coefficient relating teachers’ expectations to sophomore year achievement was .12, whereas the path coefficient relating teachers’ expectations to senior year achievement was .06. Teachers’ expectations from freshman year predicted senior year achievement less strongly than they predicted sophomore year achievement. West and Anderson (1976), however, did not assess whether .12 was significantly larger than .06. Thus, the extent to which this supports dissipation is unclear. Clearly, however, accumulation did not occur.

Smith, Jussim, and Eccles (1999)

Smith, Jussim, and Eccles (1999) examined whether teacher expectation effects accumulated from 6th through 12th grades (*N*s ranged from about 500–1700, depending on the analysis). Outcomes included both final grades and standardized test scores. All analyses controlled for students’ prior achievement test scores, grades, and motivation. The main results showed no evidence of accumulation. For the most part, expectancy effects dissipated over time. The path coefficient relating sixth grade teachers’ expectations to standardized test scores in seventh grade was .08; by 12th grade, this was .02. The path coefficient relating

seventh-grade teacher expectations to 10th-grade standardized test scores was .16; by 12th grade, it was .09.

For year-end grades, the overall expectancy effects were larger, but the declines were even steeper. The path coefficient relating sixth-grade teacher expectations to sixth-grade final marks was .33; by 12th grade, this coefficient was .17. The path coefficient relating seventh-grade teacher expectations to seventh-grade final marks was .49; by 12th grade, this coefficient was .26. Teacher perceptions in sixth and seventh grade predicted both grades and standardized test scores more weakly over time. All of these declines were statistically significant.

Although these results predominately supported the dissipation hypothesis, Smith et al. (1999) also found that the expectancy effects in 1 year were very long lasting. That is, teacher perceptions in sixth and seventh grade predicted significant changes in student achievement through high school. Thus, the durability of such effects—over many years and many different teachers—was quite striking. Durability, however, is not the same as accumulation.

Conclusions Regarding Accumulation

Concurrent accumulation most likely occurs, at least sometimes. When the path models in Figure 1 are used to interpret all existing research on teacher expectations, they show that concurrent accumulation effects in the classroom are quite small. The evidence regarding accumulation over time is even clearer. At least in the classroom, it does not happen. Four studies have directly addressed this issue (Rist, 1970; Rosenthal & Jacobson, 1968; Smith et al., 1999; West & Anderson, 1976). None found evidence of accumulation. All found at least some evidence that self-fulfilling prophecies dissipate. The three that followed students for more than 2 years, however, also found that, although self-fulfilling prophecies dissipated, they did not evaporate completely (Rist, 1970; Smith et al., 1999; West & Anderson, 1976). Thus, although there is no evidence of accumulation effects, there is good evidence that self-fulfilling prophecies that occur in 1 year can have long-lasting consequences.

Resolved and Unresolved Controversies

Can any general conclusions be reached from what may appear to be a mess of complex findings, inconsistent replications, and heated controversies? The simple answer is a clear “yes;” nearly all the controversies can be resolved by the conclusion that self-fulfilling prophecies in the classroom do exist, but they are generally small, fragile, and fleeting. This summation, however,

does not do justice to the many nuances and exceptions in the data. Although typically weak, some large self-fulfilling prophecies have been found especially regarding members of some at-risk groups; although self-fulfilling prophecies dissipate, they may endure in diluted form for years. This concluding section, therefore, summarizes how the empirical evidence bears on the six questions that framed this review, and how this review has clarified the research evidence, resolved controversies, produced useful theoretical analyses, and suggested important directions for future research.

Self-Fulfilling Prophecies in the Classroom Are Real

This review revisited and attempted to resolve some of the controversies that have plagued teacher expectation research from the start, if those controversies were still actively represented in modern scholarship. Although the conclusion that self-fulfilling prophecies are indeed real may appear to be “old news” to those long convinced, the periodic resurgence of claims denying the existence of self-fulfilling prophecies (Roth, 1995; Rowe, 1995) suggests that even this old news bears re-affirmation in a modern review.

Self-Fulfilling Effects Are Typically Small

This review has also concluded that, although self-fulfilling prophecies in the classroom are real and occasionally large, far more often, they tend to be small; the major reason teacher expectations predict student achievement is accuracy, not self-fulfilling prophecy. Although these conclusions are also old news in some circles, the periodic resurgence of claims emphasizing the power of self-fulfilling prophecies or the denial of accuracy (Claire & Fiske, 1996; Jones, 1986; Jost & Kruglanski, 2002; Schultz & Oskamp, 2000) suggests that even this old news bears re-affirmation in a modern review.

Caveats to Pygmalion

The contributions of this review, however, go beyond re-affirming old news. Although controversies surrounding Rosenthal and Jacobson's (1968) study have been well-known for years, this review has documented the frequency with which Pygmalion is still summarized in an uncritical, oversimplified manner that consistently distorts the results. The Pygmalion study has been used to justify arguments claiming that expectancy effects are powerful and pervasive, intelligence is primarily environmentally determined, and relatively simple interventions can improve student achievement. It has also been used to justify arguments emphasizing the power of beliefs to construct social re-

ality. Such uses of Pygmalion are not restricted to claims published before 1973, or even before 1993. For the many researchers who may not be aware that the entire self-fulfilling prophecy effect hinged on the occurrence of bizarre outliers and out-of-range IQ scores, the sections reviewing Snow's various critiques (Elashoff & Snow, 1971; Snow, 1969, 1995) documenting this state of affairs should constitute new, not old, news.

Many social scientists, however, may be aware of these weaknesses but choose to ignore them when discussing Pygmalion. It is, of course, a matter of scientific judgment how much of any study to believe. Therefore, this review has documented the highly limited and constrained nature of the conclusions justified on the basis of the Rosenthal and Jacobson (1968) study, even if its results are taken entirely at face value.

Putting the IQ Controversy in Perspective

This review also shows how prior reviews and meta-analyses have frequently reached seemingly diametrically opposed and mutually exclusive conclusions regarding the effect of teacher expectations on student intelligence (Raudenbush, 1984, 1994; Snow, 1995; Spitz, 1999; Wineburg, 1987). An important aspect of this contribution, however, has been to point out that, although debate between the different positions is often heated, the degree of factual disagreement between them is actually quite small. If one believes the critics, the IQ effect is zero. If one believes the advocates, it is very small (frequently 0, never consistently much higher than an r of .2). The present review has not resolved this remaining degree of disagreement. It has pointed out, however, something that may have been lost in the heat of the controversy: Although the scientific evidence may be equivocal regarding whether teacher expectation effects on IQ are nonexistent or reliably very small, it is completely unequivocal that such effects, if they occur at all, are not very large by any standard.

Positive Versus Negative Expectancies

This review has also critically evaluated the evidence regarding the relative power of positive versus negative expectancy effects. Given the frequency with which psychologists have emphasized the relatively greater power of negative expectations, this review's documentation of the paucity of evidence on this issue, and its contradictory results, is a particularly important contribution. Although strong conclusions emphasizing the inordinate power of negative expectations may someday be justified by empirical evidence, such justification will require considerably more data than is currently available.

The Nature and Extent of Accumulation

This review distinguished between two different types of accumulation—across perceivers and over time—and presented a theoretical analysis of the processes by which each type of accumulation might occur, and of the processes that might prevent accumulation. This theoretical analysis showed that concurrent accumulation cannot exceed the (typically small) coefficients representing self-fulfilling prophecies in naturalistic studies. Finally, the four studies that have addressed accumulation over time all produced essentially the same pattern of dissipation. This pattern should give any scientist considering making claims about the power of expectancy effects to accumulate considerable reason to pause.

Future Research Directions

The role of moderators. Further work on moderators of teacher expectation effects is sorely needed. With the exception of Raudenbush's (1984, 1994) meta-analyses, most work on moderators constitutes individual studies, which are bound and qualified by time, geography, subject matter, and grade levels. Research on how teacher, school, and community (urban, suburban; upper middle class, working class) characteristics moderate self-fulfilling prophecies is needed. Although, as a broad generalization, self-fulfilling prophecies are generally small, fragile, and fleeting, there is evidence that, in certain contexts and among certain groups, they are indeed consistently powerful and pervasive. Furthermore, research on moderators of the IQ effect holds out the promise of resolving the residual disagreement over whether such effects are real but small versus nonexistent.

The accuracy of teacher expectations. Just as self-fulfilling prophecies may be typically small but occasionally large, teacher expectations may be typically fairly accurate but occasionally highly inaccurate. Because little is known about the conditions under which teacher expectations become more or less accurate, research on moderators of teacher expectation accuracy is clearly important. Furthermore, although many studies and meta-analyses have addressed the extent to which teachers perceive differences between students from differing social and demographic groups (Dusek & Joseph, 1983; M. C. Taylor, 1992), the only two that have addressed whether such teacher perceptions are accurate found they were highly accurate (Jussim et al., 1996; Madon et al., 1998). Given the relevance of such research to theoretical perspectives on stereotypes and prejudice, to understanding the validity of everyday social judgment, and to assessing the role of education in creating, sustaining, or alleviating

social injustices, more work assessing this particular type and degree of accuracy is also clearly needed.

Are self-fulfilling prophecies mainly harmful or helpful? Identifying whether self-fulfilling prophecies are primarily helpful or harmful is clearly important, both regarding understanding the interpersonal dynamics between teachers and students and regarding understanding the role of expectancy effects in creating, sustaining, or alleviating social problems. As such, more research in this area is clearly needed. Furthermore, it is possible that some conditions are more conducive to negative expectancy effects, whereas others are more conducive to positive ones. Therefore, research examining such conditions holds promise for resolving the apparently conflicting results found among the only three studies to address this issue.

Do expectancy effects accumulate? Four studies demonstrating that dissipation dominates accumulation—even four studies using widely varying methodologies, conducted decades apart, and studying students at differing grade levels—is not a large body of research. Thus, more research on this issue is warranted. Furthermore, even if dissipation is the dominant phenomenon, certain conditions might be more conducive to accumulation. For example, in highly rigid, fixed systems of intergroup domination, such as those that once characterized the Jim Crow American South, South African apartheid, or the Hindu caste system, the greater distribution of resources to dominant groups may create perpetually accumulating self-fulfilling prophecies consistent with Merton's (1948) original analysis and the more modern social psychological emphasis on expectancy effects.

Conclusion

Over 30 years of active and intensive research on teacher expectations have provided important insights into basic developmental, educational, and social phenomena. Those insights, however, have sometimes become lost or distorted, at least in part, because of the vast variety of disciplines in which empirical studies, reviews, critical analyses, and meta-analyses have appeared; and because the heat of some of the controversies surrounding this research has sometimes obscured the considerable light that has also been generated. It is hoped that, by bringing this material together in a single review, the picture of what has and has not been shown by research on the self-fulfilling effects of teacher expectations now stands out much more clearly.

References

- Alwin, D., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37–47
- American Psychological Association. (1991). In the Supreme Court of the United States: Price-Waterhouse v. Ann B. Hopkins (Amicus curiae brief). *American Psychologist*, *6*, 1061–1070.
- Aronson, E. (1999). *The social animal* (8th ed.). New York: Worth.
- Babad, E., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, *74*, 459–474.
- Brattesani, K. A., Weinstein, R. S., & Marshall, H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology*, *76*, 236–247.
- Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*, 631–661.
- Brophy, J. E. (1985). Teacher-student interaction. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 303–328). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Brophy, J., & Good, T. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt.
- Claire, T., & Fiske, S. (1998). A systemic view of behavioral confirmation: Counterpoint to the individualist view. In C. Sedikides, J. Schopler, & C. A. Insko (Eds.), *Intergroup cognition and intergroup behavior* (pp. 205–231). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Darley, J. M., & Fazio, R. H. (1980). Expectancy-confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*, 867–881.
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, *56*, 225–248.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, *94*, 369–389.
- Detterman, D., & Thompson, L. A. (1997). What is so special about special education? *American Psychologist*, *52*, 1082–1090.
- Devine, P. (1995). Prejudice and outgroup perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 467–524). New York: McGraw-Hill.
- Dusek, J., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, *75*, 327–346.
- Eagly, A. H., Makhijani, M. G., Ashmore, R. D., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128.
- Elashoff, J. D., & Snow, R. E. (1971). *Pygmalion reconsidered*. Worthington, OH: Charles A. Jones.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 357–411). New York: McGraw-Hill.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*, 75–90.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Gilbert, D. T. (1995). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99–147). New York: McGraw-Hill.
- Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.
- Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues*, *46*, 35–60.
- Harris, M. J. (1991). Controversy and cumulation: Meta-analysis and research on interpersonal expectancy effects. *Personality & Social Psychology Bulletin*, *17*, 316–322.
- Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, *97*, 363–386.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*, 78–79.
- Hofer, M. A. (1994, December 26). Behind the curve. *The New York Times*, pp. A39.
- Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, *76*, 777–781.
- Humphreys, L. G., & Stubbs, J. (1977). A longitudinal analysis of teacher expectation, student expectation, and student achievement. *Journal of Educational Measurement*, *14*, 261–270.
- Ickes, W. (1997). *Empathic accuracy*. New York: Guilford.
- Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, *39*, 1–123.
- Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. Vol. 1 (4th ed., pp. 47–107). New York: Random House.
- Jones, E. E. (1986). Interpreting interpersonal behavior: The effects of expectancies. *Science*, *234*, 41–46.
- Jones, E. E. (1990). *Interpersonal perception*. New York: Freeman.
- Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*, 1–27.
- Jost, J. T., & Kruglanski, A. W. (2002). The estrangement of social constructionism and experimental social psychology: History of the rift and prospects for reconciliation. *Personality and Social Psychology Review*, *6*, 168–187.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach*. Orlando, FL: Harcourt.
- Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, *93*, 429–445.
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, *57*, 469–480.
- Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *98*, 54–73.
- 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. (1995). Naturalistic studies of interpersonal expectancies. *Review of Personality and Social Psychology*, *15*, 74–108.
- Jussim, L., Eccles, J., & Madon, S. J. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in experimental social psychology*, *29*, 281–388.
- Jussim, L., Smith, A., Madon, S., & Palumbo, P. (1998). Teacher expectations. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 7, pp. 1–48). Greenwich, CT: JAI.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308.
- Madon, S. J., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, *72*, 791–809.
- Madon, S. J., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class and ethnic

- stereotypes: Naturalistic studies in person perception. *Personality and Social Psychology Bulletin*, 24, 1304–1318.
- Madon, S. J., Smith, A., Jussim, L., Russell, D. W., Eccles, J., Palumbo, P., et al. (2001). Am I as you see me or do you see me as I am: Self-fulfilling prophecies versus self-verification. *Personality and Social Psychology Bulletin*, 27, 1214–1224.
- Mazella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*, 24, 1315–1344.
- McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85, 314–322.
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8, 193–210.
- Meyer, W. J. (1985). Summary, integration, and prospective. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 353–370). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology*, 37, 233–256.
- Myers, D. G. (1999). *Social psychology* (6th ed.). New York: McGraw-Hill.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.
- Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology*, 56, 374–386.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Oakes, J. (1986). Tracking, inequality, and the rhetoric of reform: Why schools don't change. *Journal of Education*, 168, 60–80.
- Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–239). New York: Guilford.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy inductions: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76, 85–97.
- Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 301–321). New York: Sage.
- Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40, 411–451.
- Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms*. New York: MSS Modular.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R. (1985). From unconscious experimenter bias to teacher expectancy effects. In J. Dusek (Ed.), *Teacher expectancies* (pp. 37–65). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3, 176–179.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and student intellectual development*. New York: Holt.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377–386.
- Roth, B. M. (1995, January 2). We can throw teacher expectations on the IQ scrap heap. *New York Times*, p. A25.
- Rowe, D. C. (1995, January 2). Intervention fables. *New York Times*, p. A25.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.
- Schultz, P. W., & Oskamp, S. (2000). *Social psychology: An applied perspective*. Upper Saddle River, NJ: Prentice-Hall.
- Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, 22, 93–121.
- Smith, A., Jussim, L., & Eccles, J. (1999). Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *Journal of Personality and Social Psychology*, 77, 548–565.
- Smith, A., Jussim, L., Eccles, J., Van Noy, M., Madon, S. J., & Palumbo, P. (1998). Self-fulfilling prophecies, perceptual biases, and accuracy at the individual and group level. *Journal of Experimental Social Psychology*, 34, 530–561.
- Snow, R. E. (1969). Unfinished Pygmalion. *Contemporary Psychology*, 14, 197–200.
- Snow, R. E. (1995). Pygmalion and intelligence? *Current Directions in Psychological Science*, 4, 169–171.
- Snyder, M. (1984). When belief creates reality. *Advances in experimental social psychology*, 18, 247–305.
- Snyder, M., & Stukas, Jr., A. A. (1998). Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology*, 50, 273–303.
- Spitz, H. H. (1999). Beleaguered Pygmalion: A history of the controversy over claims that teacher expectancy raises intelligence. *Intelligence*, 27, 199–234.
- Stangor, C. (1995). Content and application inaccuracy in social stereotyping. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy* (pp. 275–292). Washington, DC: American Psychological Association.
- Sutherland, A., & Goldschmid, M. L. (1974). Negative teacher expectation and IQ change in children with superior intellectual potential. *Child Development*, 45, 852–856.
- Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457–477.
- Swann, W. B., Jr. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology*, 53, 1038–1051.
- Swann, W. B., Jr., & Ely, R. J. (1984). A battle of wills: Self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology*, 46, 1287–1302.
- Sweeney, L. T., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of experimental studies. *Behavioral Science and the Law*, 10, 179–195.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay vs. John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105, 409–429.
- Taylor, M. C. (1992). Expectancies and the perpetuation of racial inequality. In P. D. Blanck (Ed.), *Interpersonal expectancies*. (pp. 88–124). New York: Cambridge University Press.
- Thorndike, R. L. (1968). Review of Pygmalion in the classroom. *American Educational Research Journal*, 5, 708–711.
- Trouilloud, D., Sarrazin, P., Martinek, T., & Guillet, E. (2002). The influence of teacher expectations on students' achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology*, 32, 1–17.
- Weinstein, R. S., & McKown, C. (1998). Expectancy effects in "context": Listening to the voices of students and teachers. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 7, pp. 215–242). Greenwich, CT: JAI.
- West, C., & Anderson, T. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research*, 46, 613–630.
- Williams, T. (1976). Teacher prophecies and the inheritance of inequality. *Sociology of Education*, 49, 223–236.
- Wineburg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy: A critical appraisal. *Educational Researcher*, 16, 28–40.