

# The Self-Fulfillment of the Self-Fulfilling Prophecy

A Critical Appraisal by SAMUEL S. WINEBURG

In the fall of 1948, a 17-page essay appeared in *The Antioch Review* that altered how we speak about social life. It did so by inventing a term to describe a widespread yet poorly understood social phenomenon. The man who coined the term was Robert K. Merton; the term was also the title of his paper, "The Self-Fulfilling Prophecy."

What began as a neologism of theoretical sociology has become common parlance. Sportswriters hurl the self-fulfilling prophecy from their columns; legislators issue it from the rostrums of Congress; and even a president has been known to invoke it, hoping thereby to find a verbal tonic for an ailing economy (Richard Nixon, 1971, quoted in Merton, 1981). Thousands of scholarly papers employing the term have appeared in sociology, social psychology, economics, political science, anthropology, public administration, and social work. But the self-fulfilling prophecy has wielded greatest influence—and doubtless stirred the most controversy—in education.

In its original form, the self-fulfilling prophecy scarcely seemed controversial. The notion that a false but widely-believed prediction could become true simply because enough people believed in it was neither new nor original. Merton, in fact, saw the term as convenient shorthand for W. J. Thomas's famous dictum that "if men define situations as real, they are real in their consequences." But the idea predated Thomas, finding roots in Bishop Bossuet's defense of Catholic orthodoxy in the 17th century; in Marx's critique of the Hegelian dialectics of change; and in Freud's work in places too numerous to count.

The self-fulfilling prophecy was a uniquely social idea with no corollary in the physical or natural sciences. Predictions of earthquakes may set Californians on edge, but words do not make the ground tremble. But in the social sphere, self-fulfilling prophecies abounded. So, for example, rumors about the insolvency of banks, when widely believed and acted upon, brought doom to otherwise flourishing financial institutions. Although the term "self-fulfilling prophecy" was new, the idea was not, and Merton wove it into his text with no claim of originality. Indeed, he abjured pretenses of innovation: "So common is the pattern of the

*Few ideas have influenced educational research and practice as much as the notion of the self-fulfilling prophecy. Yet despite the impact of the "educational self-fulfilling prophecy," the empirical status of the construct, as applied to the effect of teacher expectations on student IQ, remains equivocal. Lack of evidence was apparent even in the now-classic "Pygmalion Study" (Rosenthal & Jacobson, 1968), which claimed to show that teachers' expectations affect student intelligence. Despite criticism of the study and failure to replicate the effect in different settings, Pygmalion became one of the most cited studies in education. What made the educational self-fulfilling prophecy so durable when many other ideas in education have a fleeting half-life? How was research that was much contested in scholarly circles used by the courts and portrayed in the media? This paper examines these questions.*

self-fulfilling prophecy that each of us has his favored specimen" (p. 195).

The self-fulfilling prophecy begins, according to Merton, with the false definition of a situation, which in turn engenders behavior that brings the situation into conformity with the definition. "The Negro is a strikebreaker and a traitor to the working class," claimed union officials in barring blacks from labor unions prior to World War II. Hungry and out-of-work, thousands of blacks couldn't resist the invitation by strike-

bound employers to take jobs otherwise closed to them. Did blacks become scabs because of their "inability to participate in collective bargaining," or because, excluded from union membership, they had no choice but to accept what was offered? For Merton, the final arbiter in cases of the self-fulfilling prophecy was *time*:

History creates its own test of the theory of self-fulfilling prophecies. That Negroes were strikebreakers because they were excluded from unions (and from a large range of jobs) rather than excluded because they were strikebreakers can be seen from the virtual disappearance of Negroes as scabs in industries where they have gained admission to unions in the last decades. (p. 197)

Quoting de Tocqueville, Merton asserted that self-fulfilling prophecies were rooted in a social structure created by humans and thus open to change by them: "What we call necessary institutions are often no more than institutions to which we have grown accustomed" (p. 210).

## The Self-Fulfilling Prophecy, Desegregation, and Teachers' Expectations

America on the eve of the 20th century was a land that had grown accustomed to its social institutions, especially its system of separate schools for children of different colors. But by this century's midpoint much had changed, and

SAMUEL S. WINEBURG is at the Center for Educational Research, School of Education, Stanford University, Stanford, California 94305.

claims that the "necessary institutions" of separate schools were no more than "mere customs"—and inhumane customs at that—were heard with greater frequency. Separating children on the basis of skin color, claimed critics, flouted the ideals of a democratic society. This was not all. Segregation's effects extended beyond institutional spheres—more was at stake than inequality of social opportunity. Segregation, according to these claims, caused *psychological* damage to its victims, engendering maladaptive internal states that left the personality scarred. These were new and serious charges, and their documentation became a social priority. Society looked to the universities for evidence, and the universities looked to the social sciences.

The research program of Kenneth B. Clark stands out above the rest. In Clark's work (e.g., Clark, 1955/1963; Clark & Clark, 1939) black youngsters between the ages of 3 and 7 were presented with two dolls, one black and one white, and asked by an adult experimenter which doll they wanted to play with. At each age level, the majority of children rejected the black doll. Sometimes youngsters showed extreme reactions, as did one little girl who, after choosing the white doll, called the black one "ugly" and "dirty." This girl sobbed uncontrollably when the researcher asked her to identify herself with one of the two dolls. Clark, a black social psychologist then at the City College of New York, summarized the implications of such an episode:

As minority-group children learn the inferior status to which they are assigned and observe that they are usually segregated and isolated from the more privileged members of their society, they react with deep feelings of inferiority and with a sense of personal humiliation. . . . Like all other human beings, they require a sense of personal dignity and social support for positive self-esteem. . . . Under these conditions, minority-group children develop conflicts with regard to their feelings about themselves and about the values of the group with which they are identified. (1955/1963, p. 63)

Not fleeting or easily healed, this feeling of inferiority was "enduring or lasting as the situation endured, changing only in its form and in the way it manifested itself" (*Brown v. Board of Education*, 1954, pp. 89-90). Taught by society to be inferior, black children learned to feel and act inferior. Aided by the mechanism of the self-fulfilling prophecy, the effects of racism moved from "out there" in society to inside people's heads, and became, in Clark's terms, "embedded in the personality" (1955/1963, p. 50).

In his role as social science consultant to the NAACP in *Brown v. Board of Education* (1954), Clark was as instrumental as any other social scientist in swaying the court to dismantle this country's system of separate schools. But as many of the promised benefits of desegregation failed to materialize (cf. Cohen & Weiss, 1977; Gerard, 1983; St. John, 1975; Weinberg, 1975), the task of solving the problems of disadvantaged youth came to be seen as one that required more than mixing children of different colors in the same school. According to one explanation, it was not ghetto youth but their teachers who were the problem, and the self-fulfilling prophecy, as an explanation for differential achievement between white and black students, came to be used with ever greater frequency. Simply changing the social organization of schools would have little effect on the

achievement of minority students unless a concomitant change occurred in the minds of their teachers. Therefore, because of the "importance of the role of teachers in the developing self-image, academic aspirations and achievements of their students," Kenneth Clark (1963, p. 148) began a study of teacher attitudes in ten inner-city schools.

In a report describing this research (Clark, 1963), the term "self-fulfilling prophecy" first appeared in the educational literature. Foreshadowing controversies that would rage a few years later, the term appeared in the context of a discussion of IQ tests. The argument was straightforward. Because teachers thought minority children were dumb, they didn't waste their time on them, and teachers' expectations were later borne out by students' low test scores. Clark wrote:

If a child scores low on an intelligence test because he cannot read and then is not taught to read because he has a low score, then such a child is being imprisoned in an iron circle and becomes the victim of an *educational self-fulfilling prophecy* [*italics added*]. (p. 150)

By the late 1950s and early 1960s, this view struck a responsive chord among educational researchers and practitioners. Clark's ideas about educational self-fulfilling prophecies filtered down to schoolpeople and became part of in-service training programs. In 1957, superintendents and school board members of 14 of the largest school districts in the country convened in Atlantic City to discuss the problems of urban schools. Emerging from the deliberations of the "Great Cities School Improvement Project" was a master plan for improving schools, the first plank of which called for the "development of a program of education adapted to the needs of these children" (Marburger, 1963, p. 302).

Carl Marburger, director of Detroit's Great Cities project, outlined how his city responded to the first plank of the master plan: "One major approach is our work with teachers. Improvement of schooling depends to a great extent upon more effective teaching" (Marburger, 1963, p. 303). But effective teaching in the early 1960s meant something much different from what that term means today. At issue was not how teachers made transitions from one part of the lesson to another (Anderson, Evertson, & Brophy, 1979), or how they reviewed math homework at the beginning of each lesson (Good & Grouws, 1979). The spotlight again sought the inner recesses of the mind: "At this point . . . we conceive the formula that teacher expectations have surprising impact on pupil achievement. Indeed we might even say that teacher expectations have a similar impact on pupil intelligence scores" (Marburger, 1963, p. 306).

These were bold claims, and in making them, Marburger ignored the traditional distinction between intelligence and achievement, a tenet of educational psychology since the days of E.L. Thorndike. Achievement was customarily seen as a reflection of school learning, malleable and open to training (Franzen, 1920; Gregory, 1922). Intelligence, on the other hand, was thought partially innate and partially fixed by environmental factors during the earliest years of life. Thus, one could grant the possibility that teachers' expectations affected student achievement, because expectations guided what teachers chose to teach and consequently what students learned. But the claim that *intelligence could* be raised or lowered by teacher expectations contradicted a

long-standing research tradition that had shown IQ to be a relatively stable construct after the early years of childhood.

Lacking empirical data to array before the participants at a conference on "Education in Depressed Areas" at Columbia University in 1962, Marburger did the next best thing. He borrowed data and argued by analogy. Citing an unpublished manuscript by a professor of psychology at the University of North Dakota, Marburger described an experiment that showed that when psychology students thought their albino rats were either "maze-bright" or "maze-dull," the rats, after a week of maze-running, fulfilled their experimenters' expectations. The designer of this experiment was unfamiliar to most schoolpeople. But five years later there would scarcely be an educator who did not recognize the name of Robert Rosenthal.

### A Finding in Search of Data

By the mid-1960s, the notion that teachers' expectations for minority students caused them to do poorly in school was well-established in people's minds. But one problem remained. If children who did poorly in school were expected by their teachers to do poorly, it could always be argued that expectations were based, not on self-fulfilling prophecies, but on students' past performance. Thus viewed, the educational self-fulfilling prophecy became a chicken or egg problem, something to be puzzled over endlessly but never known with certainty. That is, unless this alternative hypothesis could be ruled out.

Robert Rosenthal, by the early 1960s, had moved from the University of North Dakota to Harvard. Rosenthal (1966) had earned a reputation for a series of ingenious experiments that demonstrated how researchers influenced the results of their seemingly dispassionate investigations. By 1963, Rosenthal had already begun to wonder if the expectancy effects he discovered with psychology students and albino rats (Rosenthal & Fode, 1963) might not also operate with doctors and their patients, psychotherapists and their clients, bosses and their employees, and teachers and their students. As Rosenthal (1985), in a retrospective essay on the course of his research program, put it, "If rats became brighter when expected to then it should not be farfetched to think that children could become brighter when expected to by their teachers" (p. 44).

Rosenthal's idea to test the self-fulfilling prophecy in a school, rather than in a hospital, a factory, or a mental health clinic, was not wedded, it seems, to any deep educational interest. Among the requests he received for a copy of an article on experimenter effects was one from Lenore Jacobson, an elementary school principal in South San Francisco. Rosenthal sent her the article along with a stack of unpublished papers and "thought no more about it" (Rosenthal, 1985, p. 44). But Lenore Jacobson did. She dashed off a second note to the Harvard psychologist and in it presented a challenge. "If you ever 'graduate' to classroom children," she wrote, "please let me know whether I can be of assistance" (p. 44). From this challenge *Pygmalion in the Classroom* was born.

Rosenthal and Jacobson's experiment was unabashedly simple, elegant, and arch. Their research site was the "Oak School," located in a low-income South San Francisco neighborhood, and, in order for the experiment to work,

the researchers had to violate the good faith of teachers. First, children in this elementary school were administered a little-known IQ test. Rosenthal described the procedure: "We had special covers printed for the test; they bore the high-sounding title 'Test of Inflected Acquisition.' The teachers were told that the testing was part of an undertaking being carried out by investigators from Harvard University" (Rosenthal & Jacobson, 1968c, p. 21). An information sheet given to the teachers explained that the "Harvard Test of Inflected Acquisition" could identify children who could be expected to "bloom," or experience an intellectual growth spurt, during the ensuing academic year. The test was actually Flanagan's Test of General Ability (TOGA), a group administered IQ test. The bloomers had been chosen by means of a table of random numbers.

The subject of the "bloomers" was brought up casually at a faculty meeting during which each teacher was given a list of about 2-7 bloomers in each classroom as identified by the "Harvard Test of Inflected Acquisition." Four months later, and again at the end of the school year, the test was readministered. The findings of the study, as reported in the full-length book about it, were striking. "We find increasing expectancy advantage as we go from the sixth to the first grade; the correlation between grade level and magnitude of expectancy advantage ( $r = -.86$ ) was significant at the .03 level" (1968a, p. 74).<sup>1</sup> Elsewhere, Rosenthal and Jacobson wrote, "The results indicated strongly that children from whom teachers expected greater intellectual gains showed such gains" (1968c, p. 22).

The finding of "an increasing expectancy advantage as we go from the sixth grade to the first grade" seemed to mean two things: First, that the treatment conferred an "advantage" upon the children who received it (the bloomers) and, second, that there was some kind of progression in the data, with the magnitude of effect decreasing linearly as children became older. But did the data support this claim? Not exactly. Table 7.1 (1968a, p. 75) showed statistically significant results (or "expectancy advantages") only for children in the first and second grades. In the other four grades, the treatment either conferred a statistically nonsignificant advantage (4th grade), no difference at all (3rd grade), or a statistically nonsignificant disadvantage (5th and 6th grades). In other words, the null hypothesis could not be ruled out in 66% of the grade levels.

Odd, then, were statements like "when teachers expected that certain children would show greater intellectual development, those children did show greater intellectual development" (Rosenthal & Jacobson, 1968a, p. 82). The lack of qualifications in the assertions was even more striking in light of the mixed results of a replication experiment in two Midwestern schools by Evans and Rosenthal (1969). Described in a footnote on Page 96 of *Pygmalion*, the results of this study showed that in these two schools, control-group girls gained about 15 points on the reasoning subscale of the IQ test, whereas the treatment-group girls gained just over 5 points, a statistically significant advantage in favor of the control-group girls.

Rosenthal and Jacobson's assertions about the effects of expectations on IQ did not go unnoticed by the research community. Scholarly reviews of the book appeared in the American Psychological Association's review journal, *Contemporary Psychology*, and in AERA's main publication, *Amer-*

ican *Educational Research Journal*. Neither took kindly to *Pygmalion*.

Writing in *Contemporary Psychology*, Richard Snow, a differential psychologist at Stanford University, pointed out that TOGA was inadequately normed for children in the youngest grades. In one classroom, a group of 19 normal children had a mean IQ score of 31 on the reasoning subscale of the test (Rosenthal & Jacobson, 1968a, p. 189). Like most IQ tests, TOGA was normed to have a mean of 100. Thus, the 19 children in the "C" track of the first grade were hardly the type of children one would expect at this run-of-the-mill school. Snow asked pointedly, "Were these children actually functioning at imbecile and low moron levels?" "More likely," he hypothesized, "the test was not functioning at this grade level. . . to obtain IQ scores as low as these, given reasonably distributed ages, raw scores would have to represent random or systematically incorrect responding" (Snow, 1969, p. 198). In other words, even if some students marked their tests at random the raw scores would have been higher than those obtained. In order to score so low, some students must have refused to put pen to paper. Snow criticized other features of the book, such as the fact that additional mental ability information (readily available at the school) was not used in the analysis, the reliance on simple gain scores "even though many mean pretest differences between treatment groups equal or exceed obtained posttest differences" (p. 198), and the use of "microscopic scales to overemphasize practically insignificant differences" (p. 199). Snow concluded: "*Pygmalion*, inadequately and prematurely reported. . . has performed a disservice to teachers and schools, to users and developers of mental tests, and perhaps worst of all, to parents and children whose newly gained expectations may not prove quite so self-fulfilling" (p. 198).

Robert Thorndike, an expert in educational and psychological testing at Columbia University and the coauthor of a widely used IQ test, was no kinder to *Pygmalion*. Like Snow's criticisms, Thorndike's had less to do with the theory behind the study or even with the design employed, than with an IQ test that produced uneven and wildly unpredictable results. Thorndike (1968) questioned the validity of a measure that would yield a mean of 150.17 among the six bloomers in class 2A, an extraordinarily high score that could only be obtained if most students received a perfect score on the subtest. But because the standard deviation for these six scores was 40 points, Thorndike wondered about the students who fell above the mean. *Pygmalion*, he wrote, "is so defective technically that one can only regret that it ever got beyond the eyes of the original investigators!" (p. 708).

Not only did *Pygmalion* get past the eyes of the original investigators but, even before the book hit the streets, headlines about it splashed over the front page of the August 14, 1967 *New York Times* (Leo, 1967). (Details of the experiment's failure to replicate, however, received a scant column inch in the continuation of the story on page 20.) Robert Rosenthal appeared on national television telling the Today Show's Barbara Walters that "teachers shouldn't be allowed to teach students who they know won't learn" (NBC, May 28, 1969). And reviews of *Pygmalion* in the media hailed it as a major contribution to understanding the problems of disadvantaged students.

Robert Coles, a Harvard psychiatrist, writing in the *New Yorker* (Coles, 1969), declared that *Pygmalion's* lesson was clear: "All sorts of young children did very much better in school than others like them, presumably because their teachers expected them to become 'bloomers,' and TOGA's putative prophecy was fulfilled so conclusively that even hard-line social scientists were startled" (p. 174). Who were these "startled" social scientists, and in what forum did they shake their heads in wonderment? Coles did not say, but it was certainly not at the 1966 American Psychological Association symposium in which Rosenthal presented the study and listened to his discussant, N. L. Gage, a Stanford University expert on teaching methods, criticize it roundly (Gage, 1966). Equally reticent was Coles on the fact that in four of six grade levels there were no statistically significant differences between the bloomers and the regular pupils; nor did he have anything to say about the failure of the experiment to replicate. Although these points eluded Coles's analysis, the opportunity to point his finger at the real culprits in the American educational system did not: "The prejudices of teachers—and the effects the prejudices have on learning—come across on almost every page of this book" (p. 175).

The reports of *Pygmalion* in the press showed how the study came to stand for whatever people wanted it to, regardless of the original research questions asked by Rosenthal and Jacobson. *Pygmalion* tested the hypothesis that teachers who believe their students are due for an intellectual growth spurt will, in fact, score higher on an IQ test than a group of comparable students for whom no expectations are held. But an article in *Time* reported that "a new book called *Pygmalion in the Classroom*" demonstrated that "many children fail to learn simply because their teachers do not expect them to" (September 20, 1968, p. 62, emphasis added). Not only was such an effect not documented, it was never even addressed in Rosenthal and Jacobson's study. *Time* went on to claim that the findings of *Pygmalion* "raise some fundamental questions about teacher training" and "cast doubt on the wisdom of assigning children to classes according to presumed ability" (p. 62).

One might think that only writers in the popular press would leap so quickly from the findings of a field-experiment to questions of reforming teacher education and the organization of schooling. But the idea that *Pygmalion* had direct implications for teacher training, compensatory education, and the ways in which this nation dealt with its poorest students was not the media's invention. Rosenthal and Jacobson themselves set the stage by forging links between *Pygmalion's* findings and questions about how to train our teachers and run our schools.

Writing in the *Scientific American* (1968c), they questioned the wisdom of federal programs that had been trying to come up with ways to overcome the educational handicaps of disadvantaged children. They argued that such programs rested on the assumption that disadvantaged children possessed some problem, or deficit, that must be remedied. Rosenthal and Jacobson asserted that such thinking was, at best, misguided:

Our experiment rested on the premise that at least some of the deficiencies—and therefore at least some of the remedies—might be in the schools, and particularly in the

attitudes of teachers toward disadvantaged children. In our experiment nothing was done directly for the child. There was no crash program to improve his reading ability, no extra time for tutoring, no program of trips to museums and art galleries. The only people affected directly were the teachers. (p. 23)

What conclusions were readers to draw from such statements? That compensatory education had been a waste? That changing teachers' expectancies might prove more cost-effective than programs sponsored under the Elementary and Secondary Education Act? "More attention in educational research should be focused on the teacher," the authors continued. And once researchers learned how the Oak School teachers were able to accomplish what they did, training packages could be designed so that "other teachers could be taught to do the same." This might even lead to the development of psychological instruments for earmarking teachers who could produce similar effects and weeding out those unable to do so. *Pygmalion* would thus play a role in the "sophisticated selection of teachers" (p. 23). Given enough time, it seemed, *Pygmalion's* findings would revolutionize American education.

### *Pygmalion's Followers*

The scholarly community witnessed the next major event in the history of the educational self-fulfilling prophecy when, in 1970, the *Harvard Educational Review* published an article by Ray Rist, a sociologist then at Washington University. The essay was entitled, "Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education." Running to 40 pages, Rist's work received more space in Volume 40 of *HER* than any other article that year, save for a study by Marshall Smith and Joan Bissell on the impact of Head Start. By 1970, the self-fulfilling prophecy had lost all vestiges of technical terminology, and Rist wove it into the early pages of his article with no citation.<sup>2</sup> Shedding quotation marks, the self-fulfilling prophecy had emerged as the common coin of educational discourse.

Rist studied a cohort of black children in an urban school from their entry into kindergarten until the beginning of second grade. Much of the article detailed the day-to-day activities in the kindergarten classroom of Mrs. Caplow, a black woman who, on the eighth day of the school year, assigned her kindergartners to three reading groups and sat them at different tables. Rist noticed a systematic pattern in the groupings. At table 1, closest to Mrs. Caplow, sat children with clean clothes, "processed" hair, and fluency in standard English. At the other two tables, and especially at table 3, the children came to class unwashed, smelling of "body odor" (p. 419), had "unprocessed" hair, and responded to the teacher's questions in black dialect. Rist described this seating arrangement, which endured in classroom groupings throughout the study, in no uncertain terms: "The group perceived as slow learners were ascribed a caste position that sought to keep them apart from the other students" (p. 444).

Mrs. Caplow's classroom organization became internalized by the youngsters, and soon they began to enact their status differences among themselves. Rist's field notes offer: "The children are rehearsing a play, Little Red Riding Hood. Pamela [table 1] tells the observer, 'The teacher gave

me the best part.' The teacher overheard this comment, smiled, and made no verbal response." On another occasion, Betty, poorly dressed, left class, and hid behind a door. In an authoritarian voice, Mrs. Caplow commanded her to return. Again, Rist's field notes observe: "When the child returns, Mrs. Caplow seizes her by the right arm, brings her over to the group, and pushes her down to the floor. Betty begins to cry" (p. 428).

The self-fulfilling prophecy was "in its final stages" by May of the school year (p. 425), catapulting some kindergartners to success while sentencing others to failure. In Rist's analysis, Mrs. Caplow was not a good-hearted, if ill-informed, black woman trying to do her best given the conditions of the urban school, but a woman who taught her students that "it was acceptable to act in an aggressive manner towards those from low-income and poorly educated backgrounds" (p. 430). Like other teachers in the school, Mrs. Caplow served as an "agent of the larger society," who vigilantly monitored the social interactions among her students "to ensure that proper 'social distance' was maintained between the various strata of the society as represented by the children" (p. 444). Not background experience, nutrition, parental involvement, or even the impoverished home and community environments of children at table 3 accounted for their lower school performance. The expectations of Mrs. Caplow and her colleagues doomed these children.<sup>3</sup>

By the final pages of Rist's analysis, Mrs. Caplow as person faded into Mrs. Caplow as symbol. Notions that schools created opportunity were naïve. Schools existed not to mitigate the effects of poverty but to reconfirm them, to shackle society's least able with "ascribed labels" that took on, through the mechanism of the self-fulfilling prophecy, "objective dimensions" (p. 431). Far from being a mirror of social inequity, the school "strongly shares in the complicity of maintaining the organizational perpetuation of poverty and unequal opportunity" (p. 447). In sum, Rist limned a portrait of schools that had gained wider and wider credence since the 1950s: Behind the failure of minority children lurked the bigotry of teachers.

### *Pygmalion in the Courts*

To policy makers, legislators, and jurists, the findings of social science constitute a well-stocked arsenal. It was only a matter of time before the self-fulfilling prophecy was being hurled back and forth in legal salvos over educational equity, desegregation, ability tracking, testing, and busing. When social science findings become embroiled in the adversarial system of the courts—a system that educates judges by persuading them—strange things happen. Two cases shed light on this process.

The public school system in the District of Columbia was on trial in *Hobson v. Hansen*. Issuing his opinion in June 1967, Judge J. Skelly Wright ordered Carl Hansen, D.C.'s Superintendent of Schools, to eliminate district "optional school zones" (which had allowed white students to cross neighborhood lines to attend predominantly white schools) and mandated busing to desegregate the schools. Further, Judge Wright ordered Hansen to dismantle ability tracks because of bias in the selection process. Throughout the five months of the trial, Judge Wright listened to pages of testimony from both sides' expert witnesses, with each side claiming a

monopoly on scientific truth. At one point, Judge Wright remarked, "The unfortunate if inevitable tendency has been to lose sight of the disadvantaged young students. . . in an overgrown garden of numbers and charts and jargon like 'standard deviation of the variable' . . . 'statistical significance' and 'Pearson product-moment correlations' " (quoted in Wolf, 1981, p. 265). Growing weary of arcane statistical analyses, Judge Wright found himself relying on a tried-and-true source of knowledge. The court, he wrote, "has been forced back to its own common sense" (quoted in Cohen & Weiss, 1977, p. 96).

Common sense goes hand in hand with the zeitgeist; today's common sense—like the right of women to vote and hold public office—is often yesterday's nonsense (cf. Geertz, 1983). Despite Judge Wright's reservations about the tangles of social science research, the idea of the educational self-fulfilling prophecy not only made sense but had the warrant of science behind it. A portion from his decision read:

Studies have found that a teacher will commonly tend to underestimate the abilities of disadvantaged children and will treat them accordingly—in the daily classroom routine, in grading, and in evaluating these students' likelihood of achieving in the future. The horrible consequence of a teacher's low expectation is that it tends to be a self-fulfilling prophecy. The unfortunate students, treated as if they were subnormal, come to accept as a fact that they *are* subnormal. (*Hobson v. Hansen*, p. 484)

To substantiate these claims, Judge Wright cited two studies: one by Clark (1963), which presented no data directly bearing on the self-fulfilling prophecy, and an edited chapter on the *Pygmalion* study (Rosenthal & Jacobson, 1968b). But unbeknownst to him, *Pygmalion* dealt with the overestimation, not underestimation, of children's abilities. Moreover, it presented no observational data of teachers and students, so there was no information on how teachers "treated" students. Further, no interviews were conducted with students to see whether they accepted their "subnormal" status. Although all of the points raised by Judge Wright may in fact be true, *Pygmalion* did not provide the evidence.

The educational self-fulfilling prophecy also played a significant role in *Bradley v. Milliken* (1971), a suit charging the Detroit Public School System with *de jure* segregation. The trial was not yet a day old when a former school board member testified that the gap in performance between black and white students could be explained by the fact that teachers "didn't have the expectations for the students that would be necessary if you were expecting the student to perform at his best" (quoted by Wolf, 1981, p. 112). The testimony of the plaintiff's expert witness on education, Dr. Robert Green, set in bold relief what can happen when social science research is used in the courts. The following excerpt is drawn from that testimony:

Q: Can you describe the experiment? I believe it was done in California, was it not, with elementary children?

Green: In a nutshell, what Rosenthal did was to allow teachers to begin to believe that one group of youngsters in the classroom [was] bright, and another in a classroom or, say, in a school, [was] not quite as bright, when, in fact, the data indicated there was very little difference between the

youngsters in terms of socio-economic status and past educational achievement. With this kind of background information, it was found that teachers who were given youngsters who were alleged to be not quite as bright, the youngsters' ability was quite different than those who were alleged to be brighter, which was reflected by actual achievement of the youngsters.

Q: In other words, the teacher was told these children were bright children, they tended to perform in that fashion?

Green: Yes. . . . Teachers were also told in this particular study that those youngsters categorized as being bright, that in mid-year. . . there would be a spurt upwards in terms of academic achievement. This did, in fact, occur with the bright youngsters, not with those. . . described as being dull. In terms of systematically observing the behavior of teachers, it was discovered that at mid-year. . . the teachers with the so-called bright youngsters spent more time with the class and presented them with more information, more educational data, and they did, in fact, cause the spurt. (quoted in Wolf, 1981, p. 112-113)

This testimony was entered into the court record without note of the inaccuracies it contained. In *Pygmalion*, teacher expectations were created only for bloomers; other students (the control group) were regarded as normal, not "less bright" and never "dull." Moreover, *Pygmalion* presented no systematic data about the socio-economic status of individuals, and there was actually a 4-point difference on the *pretest* in favor of the 1st- and 2nd-grade bloomers (a seemingly minor point that looms large when one remembers that these were the grade levels in which an effect was found).<sup>4</sup> Not only was there no "systematic" observation of teachers in *Pygmalion*, there was *no* observation. This fact made it impossible to know if teachers "spent more time" with the bloomers, "presented them with more information," or taught them "more educational data." Beliefs about *Pygmalion* had become nearly as important as its actual findings, and separating fact from fancy would prove to be no easy task.<sup>5</sup>

### *Pygmalion* Reconsidered in Educational Research

*Pygmalion's* influence on educational research is nothing short of remarkable—since the original study there have been, by one estimate (Meyer, 1985), between 300 and 400 published reports related to the educational self-fulfilling prophecy (for reviews see Brophy, 1983; Brophy & Good, 1974; Cooper & Good, 1983; Dusek, 1975, 1985). A research tradition that began by inducing teachers to form false expectations based on experimental manipulation quickly progressed to the study of expectations as they naturally occurred in ordinary classrooms. In the course of this work, researchers built and refined elaborate models for the communication of expectations (Brophy & Good, 1974; Cooper, 1979; Cooper & Good, 1983; West & Anderson, 1976) and developed finely-tuned observation systems for analyzing complex classroom interactions (Brophy & Good, 1970b).

Brophy and Good (1970a) were among the first to study naturally occurring expectations. They asked four 1st-grade teachers to rank their students in order of expected achievement and then observed high- and low-expectation students in their respective classrooms. Interesting patterns emerged. High-expectation students volunteered more answers, ini-



tiated more contacts with their teachers, raised their hands more often, and had fewer reading problems than their low expectation peers, findings which led Brophy and Good to the realization that many "teacher expectation effects" are best understood as student effects on teachers. But other findings were less easy to explain this way. For instance, when low-expectation students gave wrong answers they were less likely to receive specific feedback, and when they gave right answers they were less likely to receive praise. Although all four teachers in this study displayed relatively similar behavior patterns, a follow-up study (Evertson, Brophy, & Good, 1972, cited in Brophy & Good, 1974) yielded less uniform results. Here, only three of nine teachers resembled those in the first study; three others displayed few differences in their interactions with high- and low-expectation students; and the final three showed the opposite pattern from the first—often seeking out lows for extra help and giving them more of a chance to get the right answer. These early studies set the tone for much of the research that followed—expectations in the classroom were phenomena far more complex and multi-dimensional than most had imagined (Dusek, 1985).

As research on teacher expectancies began to accumulate, teachers looked less like the villains portrayed in the earlier studies. True, some teachers ignored information from students and hewed to rigid expectations, distributing turns unfairly in reading groups (Allington, 1980), criticizing some students more harshly than others (Brophy & Good, 1970b; Good, Sikes, & Brophy, 1973), and smiling at some while reacting coolly toward others (Babad, Inbar, & Rosenthal, 1982). Every profession includes those who do it a disservice, and teaching is no exception. But for the most part, teachers' expectations proved to be based on the best evidence in their possession (Borko, Cone, Russo, & Shavelson, 1979; Shavelson, Cadwell, & Izu, 1977), and most teachers were willing to abandon initial expectations when more dependable evidence became available. Indeed, a recent study suggests that those *least* acquainted with classroom life are often most influenced by the background information they receive about students (Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987). Experienced teachers tend to disregard such information, preferring to form judgments based on their own firsthand experience with students. Summarizing much research over the years on educational self-fulfilling prophecies, Brophy (1983) concluded:

Although there are relationships between teacher expectations, teacher-student interaction, and student achievement, most of these are more accurately construed as student effects on teachers rather than as teacher expectation effects on students. Most differential teacher expectations are accurate and reality-based, and most differential teacher interaction with students represents either appropriate, proactive response to differential student need, or at least understandable reactive response to differential student behavior... although the potential for teachers' expectations to function as self-fulfilling prophecies always exists, the extent to which they actually do so in typical classrooms is probably limited... (p. 634)

One might assume that somewhere amidst all of these subsequent studies lay the vindication of *Pygmalion*, proving once and for all that teacher expectations boost (or lower)

students' IQs. If one were to rely on Rosenthal's reading of the literature, it would seem this way. Indeed, he has claimed that in the process of examining the results of 345 studies of the interpersonal self-fulfilling prophecy "some clear conclusions have emerged. The reality of the phenomenon is beyond doubt" (Rosenthal & Rubin, 1978, p. 385). To arrive at this conclusion, Rosenthal and Rubin combined studies of the self-fulfilling prophecy in areas as diverse as animal learning, ink-blot reading, reaction times, interpersonal interviews, and classroom learning.

Within education, the issue had never been whether teachers form expectancies or whether these expectancies affect students in sundry subtle and not-so-subtle ways. Acknowledging as much in his preface to Elashoff and Snow's book-length reanalysis of the *Pygmalion* data, Gage (1971) noted that previous research had not only provided support for the existence of teacher expectancies but that expectancies influenced what "teachers try to teach and thus what students learn, how students feel about themselves, how they get along with the teacher and their fellow students" (p. 5). But regarding the dispute that has come to be known as the "Pygmalion controversy" such questions missed the point.

Obscured and long forgotten, the heart of the *Pygmalion* controversy was the bold claim that *intelligence* was affected by teacher expectations. It was this claim that was heralded by the media and influenced the Los Angeles School Board to ban IQ testing in the elementary grades (cf. McCurdy, 1969). But years of replications and follow-up studies have shown that strong claims about the relationship between expectations and intelligence were unwarranted. For example, Rosenthal and his associates tried several times to replicate the expectation-IQ linkage they reported in *Pygmalion*. Evans and Rosenthal (1969) found no significant differences in total IQ after a year between the treatment and control groups. A study by Conn, Edwards, Rosenthal, and Crowne (1968) yielded no statistically significant differences between the treatment and control groups in total IQ after 4 months (though there was a 1.46 point gain score advantage to the bloomers). After 3 semesters the effect faded in total IQ, with slight nonsignificant differences favoring the control groups. In another study, Anderson and Rosenthal (1968) manipulated the expectations of counselors at a day camp for retarded boys and administered an IQ test at the beginning and end of the eight-week experimental period. The only significant IQ change was a *decrease* on the reasoning subscale for the boys who were expected to bloom, a finding that clearly ran contrary to predicted results. In a meta-analysis based on 18 studies, Raudenbush (1984) found a small mean effect size in IQ-expectation studies ( $\Delta = .11$ ), a finding that either achieved or failed to achieve statistical significance depending on the test employed. In Smith's (1980) meta-analysis, sizeable effects were found for teacher expectancies on student achievement, class participation, and social competence. But a meta-analysis of 22 teacher expectancy/IQ studies showed that the "effect of teacher expectations on pupil IQ was quite low... Pupil intellectual ability is minimally affected by the labelling information about this intellectual potential" (p. 54).

Yet such evidence often does little to dislodge beliefs about the expectation/intelligence linkage. Thus, when *Pygmalion's*

critics (e.g., Snow, 1986) try to refute claims that the self-fulfilling prophecy is responsible for enduring intellectual gains (Jones, 1986), their letters to the editor are rejected for "lack of space." Perhaps this response is understandable given the almost mythic proportions *Pygmalion* has assumed in some college textbooks. When material from the study appeared in a chapter on "Prejudice and Discrimination" in Kenneth and Mary Gergen's *Social Psychology* (1981), the data for grades 3, 4, 5, and 6 (the grade levels in which there were no significant effects) mysteriously vanished, leaving a chart containing only the *positive* effects for grades 1 and 2. It seems that, in *Pygmalion's* old age, she doesn't develop wrinkles—but loses them.

### *Pygmalion* as a Cultural Ideal

Most new ideas in education have a pitifully short half-life, but occasionally an idea captures the popular imagination and takes hold. In one form or another, the self-fulfilling prophecy had been in the educational literature since the late 1950s, but it wasn't until *Pygmalion in the Classroom* that its popularity soared. Why?

*Pygmalion* represented not merely an idea, but an *ethos*, a uniquely American way of looking at ourselves and understanding what we saw. *Pygmalion* used empirical research to document what we believed as a people: That not our social class, our previous experiences, or even the test scores in our academic files limited our ability to bloom. An ad in *Reader's Digest* put it thus: "Self-Fulfilling Prophecy—A key to success. Actual experiments prove this mysterious force can heighten your intelligence, your competitive ability, your will to succeed. The secret: just make a prediction" (quoted in Good & Brophy, 1977, p. 383). In intelligence, as in so many other spheres of American life, the road to success was paved with the power of positive thinking.

Not only was *Pygmalion* born in the right country, but her timing was impeccable. Bewildered by a far-off war and rocked by the civil rights movement, busing, and racial disturbances at home, America readily laid blame. Teachers, sufficiently powerless and disorganized, were convenient candidates.<sup>6</sup> Graphic portrayals of teacher negligence poured from all quarters—especially from bookstores where Jonathan Kozol's *Death at an Early Age* (the 1967 National Book Award winner) and Herbert Kohl's *36 Children* (1967) became best-sellers. Reviewing *Pygmalion* for the *New York Review of Books*, Kohl (1968) claimed the book's findings were consistent with what critics of schools had long been saying. Indeed, before the nation had ever heard of the Oak School or the Harvard Test of Inflected Acquisition, the educational self-fulfilling prophecy was a *social truth*, a familiar explanation of how schools undermined the intellectual performance of the disadvantaged. By the time *Pygmalion* arrived, her numbers were of less importance than her message. Her data did not substantiate a theory as much as a theory substantiated her.

Events within the educational community also spread *Pygmalion's* popularity. By the late 1960s, Arthur Jensen's ideas about the heritability of IQ were being noted (Jensen, 1969). Jensen proposed that the average difference in "IQ scores between black and white people may be attributable as much to heredity as environment" (1973, p. 80). To these claims, *Pygmalion* provided a powerful counter, asserting

that intelligence was as much a function of our social situation as anything in our genes. For years, Arthur Jensen and Robert Rosenthal squared off in perennial, sometimes acrimonious debates (e.g., Edson, 1969; Jensen, 1969, 1980; Rosenthal, 1980, 1985; Rosenthal & Rubin, 1971). But divisiveness between two psychologists is of less interest to us here than the impressions about this conflict picked up by the popular press and the public. *Pygmalion*, matching the beliefs of jurists, journalists, and the public at large, came to symbolize the American ideal (cf. Cronbach, 1975). As the cover of the May, 1970, *Family Circle* proclaimed: "Your Child's IQ Can Be Improved—New Findings."

Educational researchers are naturally drawn to studying variables they can control, and obviously teachers' expectations are much easier to manipulate than students' social class or parents' educational attainments. In *Pygmalion's* trail, teacher expectations became an educational growth industry. The "Teacher Expectancy and Student Achievement" (TESA) training program of the Los Angeles County Unified School District (cf. Kerman, 1979; Columbus Public Schools, 1982) has emerged as one of the most widely distributed in-service programs in this country, in addition to its distribution in Western Europe, Puerto Rico, England, Panama, Australia, and Saudi Arabia (Kerman, personal communication, June 15, 1987). Workshops that bring teachers' biases to the surface and heighten their awareness of classroom behavior can only be viewed positively. But when distinctions between what expectations can and cannot do become blurred, when all differences among students are cast as a function of their teachers' expectations, a dangerous trend is set. As Fein (1971) noted of such thinking:

[It] makes educational equality seem much easier to attain. No longer are we required to worry centrally about . . . environmental disadvantage; simply create a system which can provide teachers who will say to their students, in effect, that they are getting better in every way, every day and, eureka, the gap will vanish. (p. 114)

Training programs and courses in teacher education programs notwithstanding, the painful gap in school performance between children of different colors and social classes remains. This is not to dismiss the contributions made by research on the educational self-fulfilling prophecy. But *writ large*, the attempt to solve the ills of American schools by changing the expectations of teachers diverts attention from basic social inequities by claiming that the central, if not the entire, cause of school failure rests in the minds of teachers.

The process by which schools inherit the responsibility for social inequity is not well understood. Yet one thing is certain—creating high expectations for schoolchildren costs less than building new housing or funding new jobs. The omnipotence of schooling is a compelling idea in a democracy, but sometimes popularity obscures falseness. Ironically, in the same article in which Robert Merton (1948) introduced the self-fulfilling prophecy, he expressed doubts about education's ability to solve the problems caused by it: "The appeal to 'education' as a cure-all for the most varied social problems is rooted deep in the mores of America. Yet it is nonetheless illusory for all that" (p. 197). □



<sup>1</sup>This "ecological correlation," based on group as opposed to individual data, is almost certainly not equal to its corresponding individual correlation. See Robinson (1950).

<sup>2</sup>When the self-fulfilling prophecy is cited today in the educational (e.g., Schmida, Katz, & Cohen, 1987) or psychological (e.g., Jussim, 1986) literatures, it is often Rosenthal, not Merton, who is given credit for the term.

<sup>3</sup>Early in the paper, Rist noted, "The basic position to be presented in this paper is that the development of expectations by the kindergarten teacher as to the differential academic potential and capability of any student was significantly determined by a series of subjectively interpreted attributes and characteristics of that student. The argument may be succinctly stated in five propositions. First, the kindergarten teacher possessed a roughly constructed 'ideal type' as to what characteristics were necessary for any given student to achieve 'success' both in the public school and in the larger society. These characteristics appeared to be, in significant part, related to social class criteria. Secondly, upon first meeting her students at the beginning of the school year, subjective evaluations were made of the students as to possession or absence of the desired traits necessary for anticipated 'success.' On the basis of the evaluation, the class was divided into groups expected to succeed (termed by the teacher 'fast learners') and those anticipated to fail (termed 'slow learners'). Third, differential treatment was accorded to the two groups in the classroom, with the group designated as 'fast learners' receiving the majority of the teaching time, reward-directed behavior, and attention from the teacher. Those designated as 'slow learners' were taught infrequently, subjected to more frequent control-oriented behavior, and received little if any supportive behavior from the teacher. Fourth, the interactional patterns between the teacher and the various groups in her class became rigidified, taking on caste-like characteristics, during the course of the school year, with the gap in completion of academic material between the two groups widening as the school year progressed. Fifth, a similar process occurred in later years of schooling, but the teachers no longer relied on subjectively interpreted data as the basis for ascertaining differences in students. Rather, they were able to utilize a variety of informational sources related to past performance as the basis for classroom grouping" (pp. 413-414).

<sup>4</sup>Regarding pretest scores Elashoff and Snow (1971) remarked, "The fact that Grade 1 and 2 experimental children have total scores averaging 4.9 IQ points higher, and reasoning scores averaging 13.2 IQ points higher than the control group does cast doubt on the effects of randomization" (p. 158).

<sup>5</sup>*Pygmalion* was also used by the plaintiffs in *Larry P. v. Riles* (1979), but this time the defense assembled its own team of experts, including R. L. Thorndike. Judge Peckham, perhaps because of the negative comments about the study, had little to say about the educational self-fulfilling prophecy in his decision. For an account of this case see Elliotts (1987).

<sup>6</sup>Teachers did not take all of this lying down. See Alexander Visser's letter to the editor in the June 29, 1965 *Boston Globe* (reprinted in Kozol, 1967) as well as Albert Shanker's objections to the *Pygmalion* study (Shanker, 1971, p. 13).

- Allington, R. (1980). Teacher interruption behaviors during primary-grade oral reading. *Journal of Educational Psychology*, 72, 371-377.
- Anderson, D.F., & Rosenthal, R. (1968). Some effects of interpersonal expectancy and social interaction on institutionalized retarded children. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 3, 479-480.
- Anderson, L.M., Everson, C.M., & Brophy, J.E. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal*, 79, 193-222.
- Babad, E.Y., Inbar, J., & Rosenthal, R. (1982). *Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers.* *Journal of Educational Psychology*, 74, 459-474.
- Borko, H., Cone, R., Russo, N., & Shavelson, R. (1979). Teachers'

- decision making. In P. Peterson & H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications*. Berkeley, CA: McCutchan.
- Bradley v. Milliken, 338 F. Supp. (E.D. Michigan, 1971).
- Brophy, J.E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75, 631-661.
- Brophy, J.E., & Good, T.L. (1970a). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, 61, 365-374.
- Brophy, J.E., & Good, T.L. (1970b). The Brophy-Good dyadic interaction system. In A. Simon & E. Boyer (Eds.), *Mirrors for behavior: An anthology of observation instruments continued, 1970 supplement. Volume A*. Philadelphia: Research for Better Schools.
- Brophy, J.E., & Good, T.L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Reinhart, & Winston.
- Brown v. Board of Education, 347 U.S. 483 (1954).
- Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D.C. (1987). Processing and using information about students: A study of expert, novice, and postulant teachers. *Teaching & Teacher Education*, 3, 147-157.
- Clark, K.B. (1955/1963). *Prejudice and your child*. Boston: Beacon Press.
- Clark, K.B. (1963). Educational stimulation of racially disadvantaged children. In A. H. Passow (Ed.), *Education in depressed areas*. New York: Bureau of Publications, Teachers College of Columbia University.
- Clark, K.B., & Clark, M. (1939). Development of consciousness and the emergence of racial identification in Negro children. *Journal of Social Psychology*, 10, 591-599.
- Cohen, D.K., & Weiss, J.A. (1977). Social science and social policy: Schools and race. In R.C. Rist & R.J. Anson (Eds.), *Education, social science, and the judicial process* (pp. 72-96). New York: Teachers College Press.
- Coles, R. (1969, April). What can you expect? [Review of *Pygmalion in the classroom*]. *The New Yorker*, pp. 169-177.
- Columbus Public Schools. (1982). *TESA News*, 1(1).
- Conn, L.K., Edwards, C.N., Rosenthal, R., & Crowne, D. (1968). Perception of emotion and response to teachers' expectancy by elementary school children. *Psychological Reports*, 22, 27-34.
- Cooper, H.M. (1979). *Pygmalion grows up: A model for teacher expectation communication and performance influence.* *Review of Educational Research*, 49, 389-410.
- Cooper, H.M., & Good, T.L. (1983). *Pygmalion grows up: Studies in the expectation communication process*. New York: Longman.
- Cronbach, L.J. (1975). Five decades of public controversy over mental testing. *American Psychologist*, 30, 1-14.
- Dusek, J.B. (1975). Do teachers bias children's learning? *Review of Educational Research*, 45, 661-684.
- Dusek, J.B. (1985). *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Edson, L. (1969, August 31). jensenism, n. The theory that I.Q. is largely determined by the genes. *New York Times Magazine*, pp. 10-11, 40-45.
- Elashoff, J.D. & Snow, R.E. (1971). *Pygmalion reconsidered*. Worthington, OH: Jones.
- Elliots, R. (1987). *Litigating intelligence: IQ tests, special education, and social science in the courtroom*. Dover, MA: Auburn.
- Evans, J.T., & Rosenthal, R. (1969). Interpersonal self-fulfilling prophecies: Further extrapolation from the laboratory to the classroom. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 4, 371-372.
- Fein, L.J. (1971). *The ecology of the public schools: An inquiry into community control*. New York: Pegasus.
- Franzen, R. (1920). The accomplishment quotient: A school mark in terms of individual capacity. *Teachers College Record*, 21, 432-440.
- Gage, N.L. (1966, September). Discussion of the symposium on "Teachers' expectations as an unintended determinant of pupils' intellectual reputation and competence." Paper presented on the program of Divisions 15 and 8 of the American Psychological Association, New York.
- Gage, N.L. (1971). Preface. In J.D. Elashoff & R.E. Snow (Eds.), *Pygmalion reconsidered* (pp. 4-5). Worthington, OH: Jones.
- Gerard, H.B. (1983). School desegregation: The social science role. *American Psychologist*, 38, 869-877.
- Geertz, C. (1983). Common sense as a cultural system. In C. Geertz (Ed.), *Local knowledge*. New York: Basic Books.
- Gergen, K.J., & Gergen, M.M. (1981). *Social psychology*. New York: Harcourt, Brace, & Jovanovich.

- Good, T.L., & Brophy, J.E. (1977). *Educational psychology: A realistic approach*. New York: Holt, Reinhart, & Winston.
- Good, T.L., & Grouws, D.A. (1979). The Missouri mathematics effectiveness project. *Journal of Educational Psychology*, 71, 353-362.
- Good, T.L., Sikes, J., & Brophy, J. (1973). Effects of teacher sex and student sex on classroom interaction. *Journal of Educational Psychology*, 65, 74-87.
- Gregory, C.A. (1922). *Fundamentals of educational measurement*. New York: Appleton.
- Hobson v. Hansen, 269 F. Supp. 401 (1967).
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Jensen, A.R. (1973, December). The differences are real. *Psychology Today*, 80-86.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Jones, E.E. (1986). Interpreting interpersonal behavior: The effects of expectancies. *Science*, 234, 41-46.
- Jussim, L. (1986). Self-fulfilling prophecy: A theoretical and integrative review. *Psychological Review*, 93, 429-445.
- Kerman, S. (1979). Teacher expectations and student achievement. *Phi Delta Kappan*, 60, 716-718.
- Kohl, H. (1967). 36 *Children*. New York: New American Library.
- Kohl, H. (1968, September 12). Great expectations [Review of *Pygmalion in the classroom*]. *The New York Review of Books*, p. 31.
- Kozol, J. (1967). *Death at an early age*. New York: Houghton Mifflin.
- Larry P. v. Riles, 495 F. Supp. 926 (N.D. Cal. 1979).
- Leo, J. (1967, August 8). Study indicates pupils do well when teacher is told they will. *New York Times*, pp. 1, 20.
- Marburger, C.L. (1963). Considerations for educational planning. In A. H. Passow (Ed.), *Education in depressed areas* (pp. 298-321). New York: Bureau of Publications, Teachers College of Columbia University.
- McCurdy, J. (1969, January 31). Testing of IQs in L.A. primary grades banned. *Los Angeles Times*, p. 1.
- Merton, R.K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8, 193-210.
- Merton, R.K. (1981, November). Our sociological vernacular. *Columbia: The magazine of Columbia University*.
- Meyer, W.J. (1985). Summary, integration, and prospective. In J.B. Dusek (Ed.), *Teacher Expectancies* (pp. 353-370). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Raudenbush, S.W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76, 85-97.
- Rist, R.C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40, 411-451.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1980). Error and bias in the selection of data. *The Behavioral and Brain Sciences*, 3, 352-353.
- Rosenthal, R. (1985). From unconscious experimenter bias to teacher expectancy effects. In J.B. Dusek (Ed.), *Teacher expectancies* (pp. 37-65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenthal, R., & Fode, K.L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183-189.
- Rosenthal, R., & Jacobson, L. (1968a). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart, & Winston.
- Rosenthal, R., & Jacobson, L. (1968b). Self-fulfilling prophecies in the classroom: Teachers' expectations as unintended determinants of pupils' intellectual competence. In M. Deutsch, I. Katz, & A.R. Jensen (Eds.), *Social class, race, and psychological development* (pp. 219-253). New York: Holt, Rinehart, & Winston.
- Rosenthal, R., & Jacobson, L. (1968c). Teacher expectations for the disadvantaged. *Scientific American*, 218, 19-23.
- Rosenthal, R., & Rubin, D.B. (1971). *Pygmalion* reaffirmed. In J.D. Elashoff & R.E. Snow (Eds.), *Pygmalion reconsidered* (139-155). Worthington, OH: Jones.
- Rosenthal, R., & Rubin, D.B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377-415.
- Schmid, M., Katz, Y.J., & Cohen, A. (1987). Ability grouping and students' social orientations. *Urban Education*, 21, 421-431.
- Shanker, A. (1971, May 23). False research vs. the public schools. *The New York Times*, p. 13.
- Shavelson, R., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. *American Educational Research Journal*, 14, 83-97.
- Smith, M.L. (1980). Teacher expectations. *Evaluation in Education: An International Journal*, 4, 53-55.
- Snow, R.E. (1969). Unfinished *Pygmalion* [Review of *Pygmalion in the classroom*]. *Contemporary Psychology*, 14, 197-200.
- Snow, R.E. (1986). *Letter to the editor of Science*. Unpublished manuscript, Stanford University, Stanford, CA.
- St. John, N.H. (1975). *School desegregation*. New York: John Wiley & Sons.
- Teachers: Blooming by deception. (1968, September 20). *Time*, p. 62.
- Thorndike, R.L. (1968). Review of *Pygmalion in the classroom*. *American Educational Research Journal*, 5, 708-711.
- Weinberg, M. (1975). The relationship between school desegregation and academic achievement: A review of the research. *Law and Contemporary Problems*, 39, 240-270.
- West, C., & Anderson, T. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research*, 46, 185-213.
- Wolf, E.P. (1981). *Trial and error: The Detroit school segregation case*. Detroit: Wayne State University Press.

## Pygmalion Effects:

### Existence, Magnitude, and Social Importance

A Reply to Wineburg by ROBERT ROSENTHAL, Harvard University

Important questions are raised by Samuel S. Wineburg in these pages about the empirical status of the construct of the educational self-fulfilling prophecy. More specifically, questions are raised about (a) the validity of the original experiment claiming to show that teachers' expectations affected pupils' intellectual performance, (b) the current meta-analytic evidence bearing on the tenability of the

hypothesis of these *Pygmalion* effects, and (c) the social importance of these effects, should they prove to be non-zero in magnitude. The purpose of this comment is to address each of these questions.

#### The Validity of the *Pygmalion* Experiment

In this section I examine only the best-known of the criticisms of the original

*Pygmalion* experiment conducted by Rosenthal and Jacobson (1968), all of which are alluded to in Wineburg's essay. These criticisms were selected because they are the most famous, written by highly regarded, well-qualified, and clearly talented workers.

**The Jensen critique.** In his article in the *Harvard Educational Review*, Arthur R. Jensen (1969) made three criticisms. The first was that the child, rather than the classroom, had been the unit of analysis and that if the classroom had been the unit of analysis results would have been negligible. Actually, the classroom had been employed as a unit of analysis as well as the children, with essentially the