

society. He worries that this line of analysis will deflect attention from what he believes to be structural reasons for the persisting inequality in the system.

The concern seems gratuitous. No one, so far as I am aware, who has written on the topic of the self-fulfilling prophecy would make the claim that American inequality is directly attributable to the self-fulfilling prophecy of teachers.

But Wineburg resists granting any legitimacy to the view that a linkage exists between the acts of individuals and the cumulative consequences for the society. Whether one turns to the bankers, lawyers, police, doctors, social workers, judges, accountants, or teachers of American society, the argument is that their countless thousands of in-

dividual interactions and judgments have no impact on regenerating the society in which we live. He denies the possibility that the micro acts of individuals can affect the macro functions of society.

This view of America—all superstructure or all atomistic acts of isolated individuals—is a fundamental distortion of the manner in which the society ultimately hangs together. It is the interrelationship of the two that creates, recreates, and sustains a social system. And it is within this system that teachers—as all others—have a responsibility for their actions. What teachers do bears consequences, and these consequences play out in the lives of others.

Beating down the messenger in the

hopes of discrediting the message is nothing new. Wineburg's piece is but one among many that has sought, in the guise of protecting teachers from criticism, to downplay the actual contributions that they do make. Teachers are portrayed as technicians in an impersonal world. The effect, in the end, is to demean the contributions of teachers. And we have come full circle. Those who have used the concept of the self-fulfilling prophecy to note the consequences for children really do believe that teachers count—and desperately so.

The views expressed here are those of the author and no endorsement by the U.S. General Accounting Office is intended or should be inferred.

Does Research Count in the Lives of Behavioral Scientists?

A Rejoinder by WINEBURG¹

In Genesis 29:14, Jacob awakens from the revelry of his wedding night to find himself lying next to, not Rachel, his beloved, but her older sister, Leah. Even though his father-in-law, Laban, plotted this ruse, biblical law required that Jacob wed Rachel and Leah. The patriarch resigned himself to his fate, but had to devote much effort to accommodating the distinctive likes and dislikes of his two, quite different wives.

My dilemma here bears some resemblance to Jacob's, for I must accommodate in a single response two critiques that differ radically in tone and substance. On the one hand, Rist accuses me of dredging up old news and dealing only cursorily—if at all—with the deeper implications of research on teacher expectancies. On the other hand, Rosenthal skirts the substantive issues I raise and devotes his reply to defending a study whose weaknesses, in Rist's words, are "widely understood and accepted." I hope here to address these varying viewpoints, recognizing that what is "widely understood" lies very much in the eyes of the understander.

Rosenthal says little new in his response to me. Indeed, with the exception of a single sentence, nearly half of his reply (from the beginning of the "Jensen Critique" to the end of the section entitled "The Rosenthal Critique") is taken verbatim from an article published several years ago (Rosenthal, 1985).² I agree with Rist that *Pygmalion's* problems are generally understood and accepted, and for that reason I restrict myself to a few brief comments. (The interested reader can see Elashoff and Snow's [1971] book-length reanalysis of the *Pygmalion* data, which includes a rebuttal by Rosenthal and Rubin.) Fortunately, to be understood, my comments require neither expertise in meta-analysis nor a Ph.D. in statistics.

Table 1 (from Rosenthal & Jacobson, 1968a, Table 7-1,

as shown in Thorndike, 1968) lists the main results of the *Pygmalion* study. One can plainly see the significant effects in the first two grades and nowhere else. Bear in mind, then, that the original "Pygmalion effect" of which so much has been made rests on 19 scores of 1st- and 2nd-graders. This, by itself, is not problematic. But we would expect that before social scientists launched a campaign to publicize such findings—appearing on the Today Show and offering recommendations for reforming teacher education—they would have been absolutely certain that these 19 scores were unimpeachable.

Table 2 (from Appendix Tables A-2 and A-3, Rosenthal & Jacobson, 1968a, as shown in Thorndike, 1968) shows the pretest scores on the IQ measure for the 6 different tracks in the 1st- and 2nd-grades. Note that one track (1C) had a mean reasoning IQ score of 30.79; another track (1B) had a mean of 47.19. If these were normal children in a normal school, why did they receive scores that would ordinarily classify them as "mentally retarded?" Rosenthal provided the answer: "Let us first be very clear that these low IQs reflect no clerical or scoring errors. These low IQs were earned because very few items were attempted by

TABLE 1
Gains in Total IQ in the Six Grades After One Year

Grade	Control		Experimental		
	N	Gain	N	Gain	Difference
1	48	+12.0	7	+27.4	15.4
2	47	+ 7.0	12	+16.5	9.5
3	40	+ 5.0	14	+ 5.0	0
4	49	+ 2.2	12	+ 5.6	3.4
5	26	+17.5	9	+17.4	-0.1
6	45	+10.7	11	+10.0	-0.7

many of the children" (1969, p. 690).

In the light of such information, a different picture emerges. If the change from pre- to posttest scores was due to trying more items, why did Rosenthal and Jacobson claim these children showed "greater intellectual development" (1968a, p. 83), "grew intellectually" (1968b, p. 234), exhibited "greater intellectual gains" (1968b, p. 247), and blossomed in "intellectual functioning" (1968a, pp. 81, 176). A change from pre- to posttest scores may be interpreted as "intellectual growth," but given what we know about the pretest, we could just as easily attribute it to other factors—misunderstood test instructions, uncontrolled test administration, selective teacher coaching, teacher encouragement for guessing, or even chance. Add to this the fact that Rosenthal tried—and failed—on four successive occasions to replicate the original effect (Anderson & Rosenthal, 1968; Conn, Edwards, Rosenthal, & Crowne, 1968; Evans & Rosenthal, 1969; Rosenthal, Baratz, & Hall, 1974), and we must ask, from where does he draw his confidence?

Is it from the two meta-analyses on teacher expectations and student IQ? But if this were the case, why did Smith

I fear that this reentry into the details of *Pygmalion* may have obscured the main points of my paper—points which had less to do with *Pygmalion's* findings than with the ways in which these findings were communicated to the public. The fact is that no matter where one stood on *Pygmalion*, only one side of the debate reached the press, the courts, and the public at large. *Family Circle* hailed *Pygmalion* as an inexpensive and easy alternative to the "disappointingly spotty success of programs such as Head Start" (Yuncker, 1970, p. 34).³ Robert Hutchins, writing in the *San Francisco Chronicle*, lauded *Pygmalion* for providing "the explanation of the effects of socio-economic status on schooling" (1969, p. 2). So impressed was the Los Angeles School Board by the study that they banned IQ tests largely on the basis of it. Why were Americans so willing to take *Pygmalion* on faith? As I have written, *Pygmalion's* numbers were of less importance than her message. *Pygmalion* was believed because she preached to the converted.

Rist views my paper as an attempt to "undercut the utility and credibility" of the self-fulfilling prophecy. Surely a rereading of my paper would disabuse him of this notion, for I acknowledge in three separate places the importance of expectancy effects in education. Rist, however, is right about one thing—I did leave an important issue implicit. At the root of my paper lies the question of warrant: What does it mean when scholars claim to possess evidence? Why should people listen to social scientists? For that matter, why should anyone take seriously a publication bearing the name *Educational Researcher*?

Our credibility—indeed, our very livelihood—rests on our use of evidence. When scholars claim to possess evidence, particularly when that evidence speaks to issues of social moment, they should be listened to—and listened to very carefully. We must weigh evidence and consider it from all angles. And we must strive to do the impossible: Regardless of our ideological leanings, we must apply the same standard to findings that match our convictions as to those we find abhorrent.

Near the end of his ethnographic study of three classrooms in a single Saint Louis elementary school, Rist (1970) wrote these words:

It should be apparent . . . that if one desires this society to retain its present social class configuration and the disproportional access to wealth, power, social and economic mobility, medical care, and choice of life styles, one should not disturb the methods of education as presented in this study. This contention is made because what develops a "caste" within the classrooms appears to emerge in the larger society as "class" The question may quite honestly be asked, "Given the treatment of low-income children from the beginning of their kindergarten experience, for what class strata are they being prepared other than that of the lower class?" (pp. 448-449)

Here Rist's discussion moves beyond a single teacher, or even two or three teachers in a single school. His focus becomes the school *writ large*, not a brick building in Saint Louis that served as his research site. The "present social configuration" he refers to is not the social organization of Mrs. Caplow's classroom, but the social structure of American society. When he writes "given the treatment

TABLE 2
Verbal and Reasoning IQ Pretest Scores in the 1st- and 2nd-grade Classrooms

Class	N	Experimental		Control	
		Mean Verbal IQ	Mean Reasoning IQ	Mean Verbal IQ	Mean Reasoning IQ
1A	3	102.00	84.67	19	119.47
1B	4	116.25	54.00	16	104.25
1C	2	67.50	53.50	19	95.68
2A	6	114.33	112.50	19	111.53
2B	3	103.67	102.33	16	96.50
2C	5	90.20	77.40	14	82.21

(whom Rosenthal cites) reach the *opposite* conclusion, namely, that students' intellectual ability is "minimally affected" by expectancy manipulations (1980, p. 54)? With respect to Raudenbush's meta-analysis, Rosenthal states: "the effects of teacher expectations were significant for his full set of 18 studies." But are the effects as unambiguous as Rosenthal would have us believe? Raudenbush employed four different statistical tests; three showed significance. However, the fourth "indicated no effect of expectancy." What distinguished this fourth method from the other three? Raudenbush explained: "Method 4 . . . weights studies by sample sizes . . . This result gave evidence that larger sample studies tended to produce smaller effects" (1984, p. 90). Finally, one last methodological point. The mean effect size for the expectancy-IQ linkage as reported by Raudenbush is .11 (SD = .20), but as any introductory statistics student knows, the mean is notoriously sensitive to extreme values when a distribution is skewed. The *median* effect size of Raudenbush's 18 studies is but .035; ten studies yielded positive difference and eight yielded negative differences (see figure 1). What kind of phenomenon is it when nearly half the attempts to produce it yield results in the wrong direction?

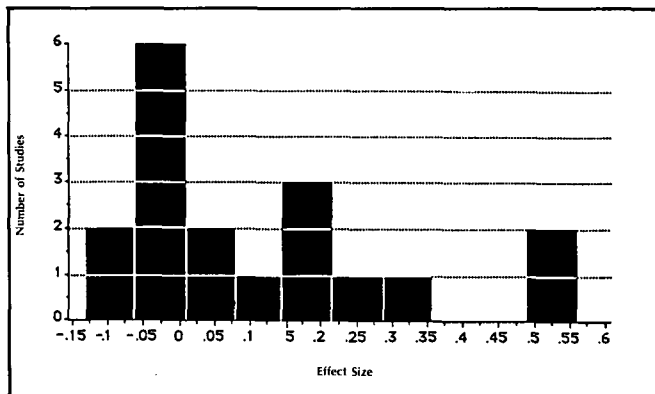
of low-income children," his discussion goes beyond a group of poor black children in a Saint Louis school to embrace the disadvantaged everywhere. When he invokes "caste," he generalizes to a nation.

Is this a credible account? Did a million American teachers in 1970, or at any time, march lockstep to the drumbeat of let's keep the downtrodden down? Were they all part of a conscious or unconscious effort to, as Rist put it, maintain "the organizational perpetuation of poverty and unequal opportunity" (1970, p. 447). I think not.

Rist began his reply by asking, "Do teachers count in the lives of students?" My answer is yes—resoundingly so—but not necessarily in the way Rist imagines. I close with a different question: Does research count in the lives of behavioral scientists, teachers, and children? If not, we might as well close up shop and refer all correspondence to *Family Circle*.

FIGURE 1

Distribution of Effect Sizes Based on Data Provided by Raudenbush (1984) (The skewness, or degree of asymmetry, of this distribution is 1.10, computed using formula 6.7 in Glass & Hopkins, 1984.)



1. At various stages in this project, I benefitted from the encouragement and admonishment of N.L. Gage, Richard E. Snow, Lee S. Shulman, Susan R. Monas, and John W. Meyer. I wish to thank them and absolve them of responsibility for any infelicities that remain. I, alone, am responsible for the views expressed here.

2. I responded to a draft copy of Rosenthal's reply that did not show ER copyediting.

3. Yuncker may have gotten this idea from statements by Rosenthal and Jacobson such as the following: "We will want to know . . . whether the efficacy of an educational practice is greater than that of the easily and inexpensively manipulatable expectation of the teacher. Most educational practices are more expensive in time and money than giving teachers names of children 'who will show unusual intellectual development' " (1968b, p. 249).

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.

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.

Elashoff, J. D., & Snow, R. E. (1971). *Pygmalion reconsidered*. Worthington, OH: Jones.

Evans, J., & 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.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Hutchins, R. (1969, August 11). Success in schools. *The San Francisco Chronicle Sunday Punch*, p. 2.

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.

Rosenthal, R. (1969). Empirical vs. decreed validation of clocks and tests. *American Educational Research Journal*, 6, 689-691.

Rosenthal, R. (1985). From unconscious experimenter bias to teacher expectancy effects. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 37-65). Hillsdale, NJ: Erlbaum.

Rosenthal, R., & Jacobson, L. (1968a). *Pygmalion in the classroom*. New York: Holt, Rinehart, and 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*. New York: Holt, Rinehart, and Winston.

Rosenthal, R., Baratz, S. S., Hall, C. M. (1974). Teacher behavior, teacher expectations, and gains in pupils' rated creativity. *Journal of Genetic Psychology*, 124, 115-121.

Smith, M. L. (1980). Teacher expectations. *Evaluation in Education: An International Journal*, 4, 53-55.

Thorndike, R. L. (1968). Review of *Pygmalion in the classroom*. *American Educational Research Journal*, 5, 708-711.

Yuncker, B. (1970, May). The power of positive thinking. *Family Circle*, pp. 34, 96, 98.