

- 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

same results. The fact had, of course, been reported in *Pygmalion* (p. 95, *passim*).

Jensen's second criticism was that the pre- and posttest of IQ had been the same so that practice effects were maximized. Jensen did not tell how practice effects could bias the results of a randomized experiment. If practice effects were so great as to drive everyone's performance up to the upper limit or ceiling of the test, then practice effects could operate to *diminish* the effects of the experimental manipulation, but they could not operate to increase those effects.

Jensen's third criticism was that teachers themselves administered the group tests of IQ and that group tests are less reliable than individually administered tests. This was a two-pronged criticism, so two points must be made. First, in the *Pygmalion* study, when children were retested by testers who were blind to the experimental conditions, indeed to the existence of experimental conditions, the effects of teacher expectations actually *increased* rather than decreased. Second, Jensen implied that the unreliability of our group test might be responsible for our teacher expectancy effects. In fact, however, lower test reliability makes it harder, not easier, to obtain significant differences between conditions. Lowered reliability decreases power, increases Type II errors, and decreases Type I errors. The basic facts about reliability have long been known (e.g., Guilford, 1954), and practical guides to usage are readily available (e.g., Rosenthal, 1984, 1987).

The Thorndike critique. A second critique of the *Pygmalion* experiment was by Robert L. Thorndike (1968). The general point was that the IQ of the youngest children was badly measured by the test employed and that, therefore, any inferences based on it must be invalid. Because a detailed discussion is available elsewhere (Rosenthal, 1969a), we give only the basic facts here.

First, the validity coefficient of the reasoning IQ subtest regarded as worthless by Thorndike was .65, quite a bit higher than the coefficients often reported in support of the validity of IQ tests. Second, even if the IQ measure had been seriously unreliable, that could not have accounted for the significant results of *Pygmalion*. As we saw

in our discussion of the Jensen critique, low reliability as a cause of spuriously significant results is a statistical nonsequitur. Low reliability can lead to fewer significant results; it cannot lead to more significant results. Third, even if the reasoning IQ data for the youngest children, the results most suspect, had been omitted from the analysis, there would still have been a significant effect of teacher expectations for the remaining classrooms as measured by reasoning IQ ($p = .001$).

The Elashoff and Snow critique. The most ambitious critique of the *Pygmalion* experiment was published as a book by Janet D. Elashoff and Richard E. Snow (1971) and discussed in detail by Rosenthal and Rubin (1971). Briefly, Elashoff and Snow transformed the original data of *Pygmalion* in eight different ways. Some of these transformations were seriously biased (e.g., discarded data unfavorable to the null hypothesis). Despite this, however, none of the transformations gave noticeably different results from those reported in *Pygmalion*. For total IQ, every transformation gave a significant result when one had been reported originally. When verbal and reasoning IQ were also considered, the various transformations yielded *more* significant teacher expectancy effects than had been reported in *Pygmalion*. Despite these results, Elashoff and Snow concluded that the *Pygmalion* research did not demonstrate teacher expectancy effects.

The new Jensen critique. The criticisms of *Pygmalion* described so far were made shortly after the appearance of that research in the late 1960s. In a more recent review, Jensen (1980) again concluded that the *Pygmalion* effect was a myth. One experiment was admitted as in support of the hypothesis of interpersonal expectancy effects. However, it was not viewed as supporting the *Pygmalion* hypothesis because it employed as a dependent variable an achievement test rather than an IQ test. Yet, three different studies showing no effect of teacher expectations were counted as evidence *against* the *Pygmalion* hypothesis, although they too employed only achievement tests as dependent variables rather than IQ tests (Rosenthal, 1980). In short, when achievement test results favored the hypothesis, they were excluded from

evidence. When they went against the hypothesis, they were included.

Also singled out for comment was a study by Deitz and Purkey (1969) "as it revealed no expectancy effect based on pupil's race" (Jensen, 1980, p. 608). Indeed it did not, because that study was not a study of teacher expectancy effects at all! In that study, rather than manipulate teachers' expectations to determine the effects on pupils' performance, the investigators asked teachers to estimate the future academic performance of black or white boys. The finding of a nonsignificant relationship between children's race and teachers' estimates of future academic success was interpreted as a failure to find an effect of teachers' expectation on pupils' IQ. This study had nothing to do with either IQ or achievement. Teacher expectation was not an independent variable at all but a dependent variable.

Of a total of 13 studies listed by Jensen as showing no "effects of teacher expectancy on children's IQs," (p. 608) four (or 31%) did not even employ IQ tests as dependent variables, and one (or 8%) did not even employ teacher expectations as an independent variable. We expect a given degree of error in science (Rosenthal, 1966, 1978), but Jensen's rate of making errors (31%) and his rate of making those errors in the direction of his hypothesis (100%) seems excessive. (Further evidence of biased reporting by Jensen is given in Rosenthal, 1980.)

Jensen also reported several studies of the expectancy effects of the examiner during the course of psychological testing. The study he singled out (Samuel, 1977) as most powerful and most informative was one that he felt provided no support for the expectancy hypothesis; "only" 6.4% of the variance was attributable to expectancy effects. The error here was in thinking that 6.4% of the variance was of little practical consequence. As Rosenthal and Rubin have pointed out (1979, 1982) and as we explain later here, accounting for 6.4% of the variance is equivalent to increasing the success rate of a new treatment procedure from 37% to 63%, a change that can hardly be considered trivial.

The Rosenthal critique. Now that we have had a look at some of the best known criticisms of *Pygmalion*, let me level two criticisms of my own. My ma-

for criticism is that, for the most part, significance tests were employed unaccompanied by estimates of the sizes of the effects discussed. I now regard that as an unacceptable mode of reporting results of data analyses (Rosenthal, 1987; Rosenthal & Rosnow, 1984, 1985). My other criticism, a more technical one, is that there were instances in which omnibus *F* tests were performed (i.e., where *df* in the numerator were greater than one). If those *F*'s had been computed at all, and there is a question as to whether they should have been, they should at least have been followed up by the linear trend contrasts (and perhaps by the quadratic trend contrasts) implied by the structure of the data. Thus, grade level (*df* = 5) and ability track level (*df* = 2) imply an interest at least in the linear effects of grade and of ability track and their interaction with other factors such as experimental condition.

The Meta-Analytic Evidence for Pygmalion Effects

There are now well over 400 experiments investigating the self-fulfilling nature of interpersonal expectations, and meta-analyses of all the studies available have appeared regularly since the late 1960s (Rosenthal, 1968, 1969b, 1974, 1985; Rosenthal & Rubin 1971, 1978). The combined probability from these meta-analyses makes the plausibility of the null hypothesis vanishingly small ($Z > 25.0$). Recent meta-analyses are also available for just the effects of interpersonal expectations on pupils' IQ test performance (Raudenbush, 1984; Smith, 1980). Both Raudenbush and Smith found significant overall effects of interpersonal expectations on students' IQ. The mean effect size reported by Smith in units of *d* [(*M1* - *M2*)/*S*] was .16; in units of *r* (the point-biserial correlation between teacher expectancy and pupil IQ), it was .08. (I postpone for the moment a consideration of the practical importance of that effect size.)

Raudenbush's more recent meta-analysis (1984) was designed to investigate the relationship between the credibility of the expectancy induction and the magnitude of the teacher expectancy effect on pupil IQ. He reasoned that inductions of expectations in teachers would be credible only to the extent that teachers did not already know the children and, thus, had not already

TABLE 1
Median Test on Raudenbush's 18 Studies Showing Effect Size as a Function of Length of Prior Teacher-Pupil Acquaintance

| Prior acquaintance | Effect size | | Σ | Percentage above median |
|---------------------|--------------|--------------|----|-------------------------|
| | Above median | Below median | | |
| Two weeks or less | 8 | 2 | 10 | 80% |
| More than two weeks | 1 | 7 | 8 | 12½% |
| Σ | 9 | 9 | 18 | |

Note. $Z = 2.85$, $p = .0022$, $r = .67$.

established expectations on the basis of their direct experience. Although the effects of teacher expectations were significant for his full set of 18 studies (as indicated by the ensemble-adjusted $p = .01$), he found dramatic differences in effect sizes as a function of how long teachers had known pupils before the induction of the expectation. (For a discussion of ensemble-adjusted p 's, see Rosenthal & Rubin, 1983.) Table 1 shows the results of a simple median test; 80% of the studies in which teachers knew pupils only two weeks or less showed effect sizes (in the predicted direction) above the median compared to only 12.5% of the studies in which teachers knew pupils longer than two weeks ($p = .0022$, $r = .67$).

The original *Pygmalion* experiment was included in Raudenbush's sample of seven studies in which teachers knew each other for one week or less. The effect size Raudenbush computed for the *Pygmalion* study was $d = .21$; the effect size he computed for the remaining six studies in this category was $d = .30$, or $r = .15$. In short, the results of the *Pygmalion* study not only were consistent with the results of all studies in that category—the results tended to underestimate the meta-analytic mean effect size.

It seems clear, then, based on the meta-analytic evidence, as well as on the evidence provided by the original *Pygmalion* experiment, that the educational self-fulfilling prophecy (Merton, 1948) has now been well established—and that is the first step in the scientific study of any phenomenon (Merton, 1987).

The Social Importance of Pygmalion Effects

For the seven experiments in Raudenbush's sample in which teachers' expectations could most credibly be created, those in which teachers knew their pupils one week or less, the mean effect sizes were $d = .29$, or $r = .14$, or $r^2 = .02$ (Raudenbush, 1984). In use, all three of these effect size estimates suffer from a common problem, the tendency to underestimate the practical importance of the effects of behavioral or biomedical interventions. Rubin and I (1979, 1982) found this tendency to underestimate the importance of effects not only among experienced behavioral researchers but among experienced statisticians as well. Accordingly, we proposed an intuitively appealing general purpose effect size display whose interpretation is far more transparent: *The Binomial Effect Size Display (BESD)*. The question addressed by BESD is: What is the effect on the success rate (e.g., survival rate, cure rate, improvement rate, selection rate, etc.) of the institution of a new treatment procedure? It therefore displays the change in outcome attributable to the new treatment procedure. (The details are given in Rosenthal & Rubin, 1982.) Here we simply illustrate the BESD for the mean effect size under discussion ($r = .14$) and for the mean effect size for just those studies Raudenbush (1984) reports in which teachers had already known their pupils for two weeks ($r = .04$).

Table 2 shows the BESD's for r 's of .14 and .04. Obviously, a treatment that changes success rates (e.g., survival)

TABLE 2
Binomial Effect Size Displays (BESD) for r 's of .14 and .04
that Account for "Only" 2% and "0%" of the Variance, Respectively

| Effect size | Condition | Example of a treatment result | | Σ |
|------------------------------|-----------|-------------------------------|------|----------|
| | | Alive | Dead | |
| $r = .14$ ($r^2 = .02$) | Treatment | 57 | 43 | 100 |
| | Control | 43 | 57 | 100 |
| | Σ | 100 | 100 | 200 |
| $r = .04$ ($r^2 = .00$) | Treatment | 52 | 48 | 100 |
| | Control | 48 | 52 | 100 |
| | Σ | 100 | 100 | 200 |

from 43% to 57% or even from 48% to 52% is of substantial practical importance. Death rates are dramatic but so are failure rates, rejection rates, and error rates. To employ the BESD, we compute the effect size r for our experimental effect. Adding the quantity $r/2$ to .50 gives the experimental group success rate. Subtracting the quantity $r/2$ from .50 gives the control group success rate. When effect sizes are displayed as a BESD, we get a better picture of the real-world importance of any treatment effect. Even so "small" an effect as one accounting for nearly 0% of the variance, the BESD shows, has practical implications to a degree that most behavioral researchers and most statisticians find surprising. If it seems far-fetched to consider important those effects improving the rate of favorable outcomes from 48% to 52%, it should be noted that a major drug study was discontinued because the effect size, which was of just that magnitude, was regarded as so important that it seemed unethical to deprive the placebo control patients of the benefits of the new drug, propranolol (Kolata, 1981).

Conclusion

Significance testing and effect size estimation based on meta-analytic procedures since the original *Pygmalion* study show not only that there is a phenomenon to be explained (Merton, 1987) but also that the phenomenon is nontrivial in magnitude. Still, there is much to learn. One area of "specified ignorance," as Merton (1987) referred to such domains, has been just how *Pygmalion* effects are transmitted. As Wineburg notes, the original *Pygmalion* study did not include observation of

teachers' differential treatment of children for whom they had or had not been given expectations for favorable intellectual development. Since that study of 20 years ago, however, scores of studies have made such observations, and 31 meta-analyses on these studies have been reported (Harris & Rosenthal, 1985). These meta-analyses gave strong support to a four factor "theory" of the mediation of interpersonal expectancy effects (Rosenthal, 1974, 1981).

Future research on teacher selection and teacher training that takes into account the effects of educational self-fulfilling prophecies may help to improve both selection and training, but for now the evidence in both domains is preliminary at best. Policy implications of the *Pygmalion* research broadly defined are, therefore, not clearly etched and remain more a subject for essays than for data-based meta-analytic work. One policy inference I am willing to draw even in the absence of such work, however, is frankly personalized: When our children were in school we didn't want them taught by teachers who "knew" they couldn't learn. Now that our children are grown, we don't want our grandchildren taught by teachers who "know" they can't learn. Sometimes assigning some Bayesian priors seems simply prudent.

Acknowledgement: Preparation of this paper and much of the research cited was supported by the National Science Foundation.

Deitz, S. M., & Purkey, W. W. (1969). Teacher expectation of performance based on race of student. *Psychological Reports*, 24, 694.
 Elashoff, J. D., & Snow, R. E. (1971). *Pygmalion reconsidered*. Worthington, OH: Charles A. Jones Publishing.

- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Kolata, G. B. (1981). Drug found to help heart attack survivors. *Science*, 214, 774-775.
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8, 193-210.
- Merton, R. K. (1987). Three fragments from a sociologist's notebooks: Establishing the phenomenon, specified ignorance, and strategic research materials. *Annual Review of Sociology*, 13, 1-28.
- 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.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1968). Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. *Psychological Bulletin Monograph Supplement*, 70 (6, Pt. 2), 30-47.
- Rosenthal, R. (1969a). Empirical vs. decreed validation of clocks and tests. *American Educational Research Journal*, 6, 689-691.
- Rosenthal, R. (1969b). Interpersonal expectations. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 181-277). New York: Academic Press.
- Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms*. MSS Modular Publications, New York, Module 53, pp. 1-28.
- Rosenthal, R. (1978). How often are our numbers wrong? *American Psychologist*, 33, 1005-1008.
- Rosenthal, R. (1980). Error and bias in the selection of data. *The Behavioral and Brain Sciences*, 3, 352-353.
- Rosenthal, R. (1981). Pavlov's mice, Pfungst's horse, and Pygmalion's PONS: Some models for the study of interpersonal expectancy effects. In T. A. Sebeok & R. Rosenthal (Eds.), *The Clever Hans phenomenon* (pp. 182-198). Annals of the New York Academy of Sciences, No. 364.
- 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. G. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 37-65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis*. New York: Cambridge University Press.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupil's intellectual development*. New York: Holt, Rinehart & Winston.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data*

- analysis. New York: McGraw-Hill.
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. New York: Cambridge University Press.
- Rosenthal, R., & Rubin, D. B. (1971). Pygmalion reaffirmed. In J. D. Elashoff and R. E. Snow, *Pygmalion reconsidered* (pp. 139-155). Worthington, OH: C. A. Jones.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377-386.
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, 9, 395-396.
- Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Rosenthal, R., & Rubin, D. B. (1983). Ensemble-adjusted *p* values. *Psychological Bulletin*, 94, 540-541.
- Samuel, W. (1977). Observed IQ as a function of test atmosphere, tester expectation, and race of tester: A replication for female subjects. *Journal of Educational Psychology*, 69, 593-604.
- Smith, M. L. (1980). Teacher expectations. *Evaluation in Education*, 4, 53-55.
- Thorndike, R. L. (1968). Review of *Pygmalion in the classroom*. *American Educational Research Journal*, 5, 708-711.
- Wineburg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy: A critical appraisal. *The Educational Researcher*, 16 (9).

Do Teachers Count in the Lives of Children?

A Reply to Wineburg by
RAY C. RIST, U.S. General Accounting Office

I have a strong sense of what Wineburg's article is and what it is not. It is a polemical piece that uses intellectual shortcuts and selective presentation of material to support the author's view regarding the use and abuse of a particular theoretical concept—that of the self-fulfilling prophecy. It is not a tightly reasoned critique or synthesis of a body of research. Though he castigates others for these faults, the author is clearly guilty of excesses of language and restatement of positions to score points. Nonetheless, the piece is provocative. As one of the outside "blind reviewers" asked by the *Educational Researcher* to assess the article, I did recommend publication.

I thought when I first read the paper (and still do) that it merited consideration in the educational research community because of two deeply important philosophical issues embedded within it. One of the issues is addressed explicitly by the author, and one remains implicit. To take the latter first: Central to the author's critique is the issue of whether an intellectual construct can have a social or cultural validity while empirical efforts to measure it remain weak and even misguided. What seems to disturb author Wineburg is that the intellectual vitality and applications of the concept have far outpaced efforts to test, refine, and

bound it. Stated differently, he seems to reject the presumption that cultural realities need not coincide with statistical realities.

Another concern raised in this paper—explicitly in this instance—merits attention. The author questions whether consequences exist for students owing to teachers' differential expectations. The argument throughout the paper is that the effects are few if any—presumably because the tests and measures to demonstrate them are unpersuasive. But, as I noted above, the piece does not seriously address research findings. Instead, it simply discredits the self-fulfilling prophecy concept, seeing it as a tool for "teacher bashing." The intent is to undercut the utility and credibility of a concept that provides an explanation of how teachers differentially set expectations for students and then act upon them. The impulse is to protect teachers from the accusations that their behaviors have consequences—both positive and negative—in the lives of children.

Although the impulse may be defensible, the author never does actually refute the proposition that teacher expectations do have outcomes for children. Instead, he hammers on the *Pygmalion* study to point up its widely understood and accepted weaknesses. The deeper questions are not addressed

and remain obscured in the rhetoric of failed research designs and unsubstantiated hypotheses.

One of the most interesting intellectual ploys in this paper is the author's emphasis upon linking teacher expectations to student intelligence. Although this link, indeed, was made by Rosenthal and Jacobson in their book, it sets the debate here in the narrowest of parameters. In this construction of the issue, the author sets out to convince the reader that something is amiss. He avoids the fact that the vast majority of work on the self-fulfilling prophecy never has been interested in linking teacher expectations to student intelligence, but rather to the consequences that teacher expectations have for the school experiences of students. The only attention given by the author to this point comes in footnote #3 where he quotes from a description I gave nearly 18 years ago in the *Harvard Educational Review* of how to link expectations and classroom consequences.

It is one thing to refute an argument by showing that a claimed outcome is not well conceptualized, let alone well measured—in this instance, the particularly difficult linkage of expectations and intelligence. It is quite another to ignore the body of work that has demonstrated changes in classroom relations, teacher-student interactions, and students' behavior toward one another when they knew who received consistently positive or consistently negative attention from teachers. To stress changes in IQ (and all the mystical presumptions surrounding such changes) as the bottom line test of whether the concept has any empirical foundation is fundamentally to miss the point, with no apologies to Rosenthal and Jacobson.

Indeed, the author himself appears to accept the vitality of the concept in his opening remarks and again within some of the concluding comments. (And this is not to ignore the presumably careful way in which the words were chosen for the title.) So, if the concept itself is not the issue, then we have to return to the basic question of what has disturbed Wineburg and thus has prompted this paper. The answer is to be found in the next to last paragraph of the paper. The author believes that the self-fulfilling prophecy should not be used as a rationalization for existing inequalities in American