EDUCATION

Pygmalion and Intelligence?

Richard E. Snow

Rosenthal has again presented his *Pygmalion in the Classroom* study as the best example of interpersonal expectancy effects, cloaking it for support in meta-analyses of related though not comparable research.[1] In my view, both his data and his field of work are thus misrepresented to both psychological science and the public. The dispute about *Pygmalion* has a long history with many facets.[2] Without trying to rehash it all in a small space, I focus here on the core malady and a suggested cure.

To be clear, I agree that the general evidence shows that interpersonal expectancies exist as psychological phenomena, and that teacher expectancies, as an example, can influence classroom teaching and learning, at least sometimes.[3] But I do not agree that the evidence shows an influence of teacher expectancy on learner intelligence.

THE MALADY

Two problems are compounded here. One arises from flaws in the original *Pygmalion* intelligence data, the other when these and diverse other data are combined in meta-analyses indiscriminately.

The original *Pygmalion* study compared experimental children, for whom positive expectancies had been suggested to teachers, with control children, for whom no expectancies had been suggested. The children were students in six grades in one elementary school. The finding was that positive experimental-control differences in total IQ scores occurred in Grades 1 and 2, not in Grades 3 through 6, and came mainly from the reasoning subscores, not the verbal subscores. Figure 1 shows the scatterplot and regressions of posttest on pretest reasoning subscores for first and second graders, separately for experimental and control groups. It also shows a box delimiting the test's norm range (IQ from 60 to 160), with separate experimental and control regressions computed within it.

About 35% of the scores fall outside the norm range. The test's scale of measurement is extrapolated out-

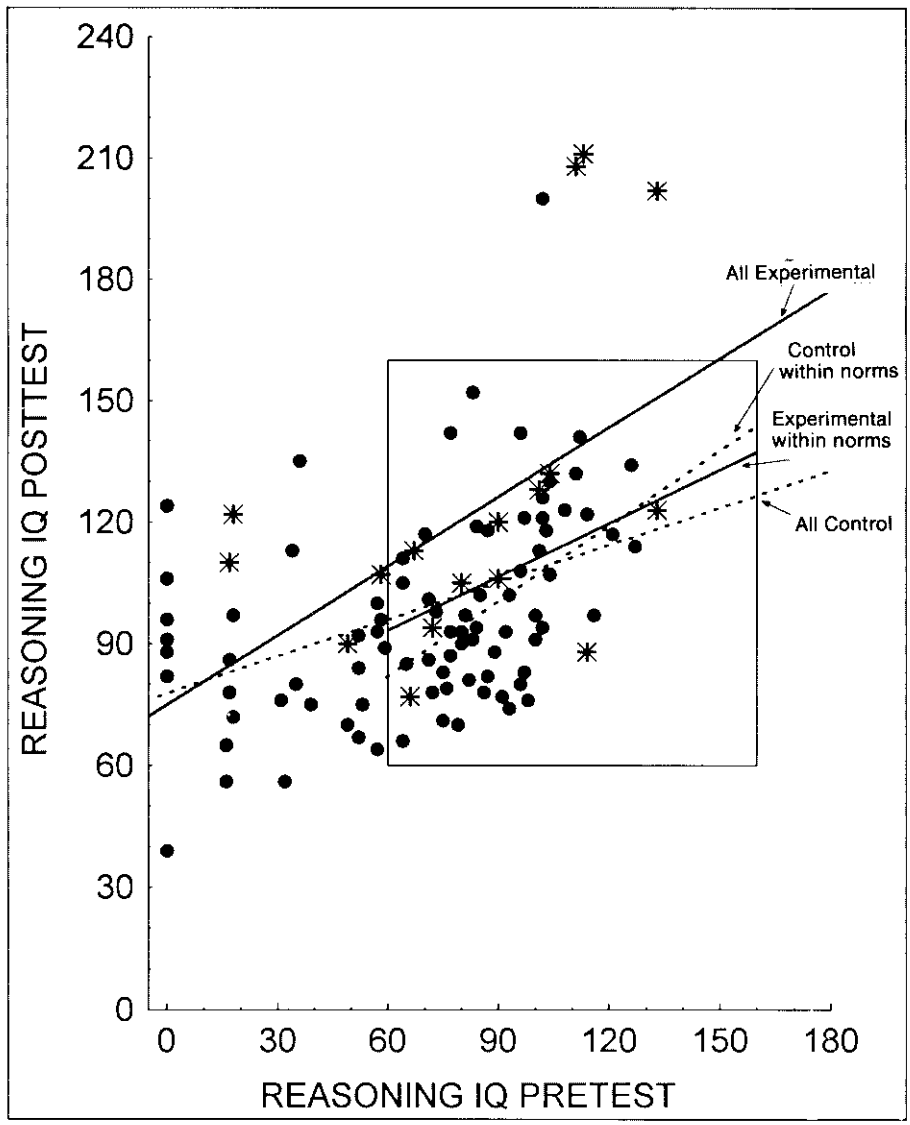

Fig. 1. Scatterplot and regressions of posttest on pretest reasoning IQ for Grades 1 and 2 of the *Pygmalion* study.[1] Separate regression lines are shown for all experimental children (the asterisks), all control children (the black dots), and those experimental and control children with scores within the norm range (i.e., within the box). I thank Robert Rosenthal for graciously providing the original data.

Richard E. Snow is Professor of Education and Psychology at Stanford University. Address correspondence to him at School of Education, Stanford University, Stanford, CA 94305-3096.

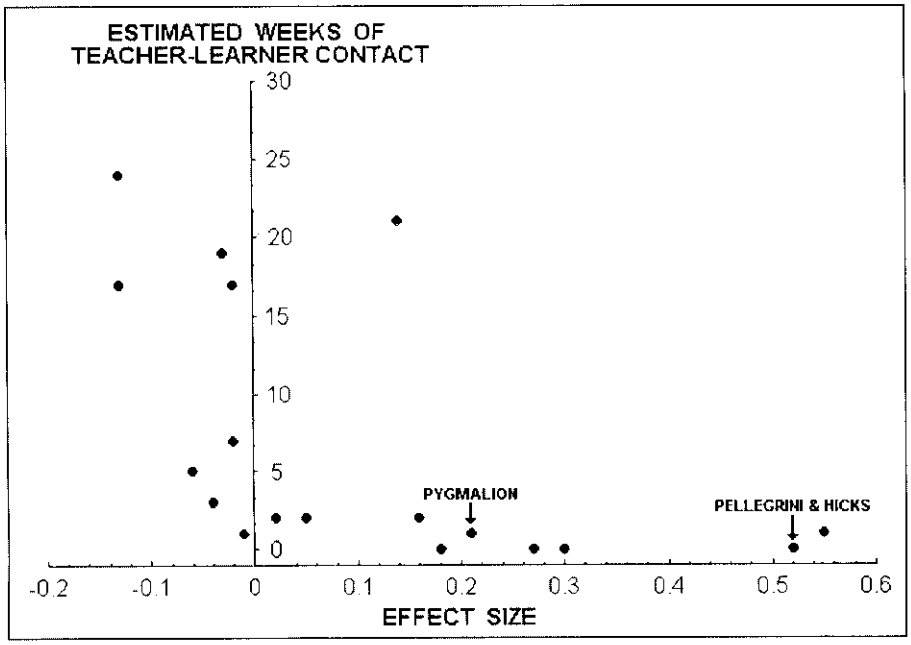

Fig. 2. Effect sizes for 18 studies plotted by estimated weeks of teacher-learner contact as reported by Raudenbush.[7] The original *Pygmalion* study[1] and the Pellegrini-Hicks study[8] are identified.

side this range and not meaningfully interpreted as intelligence. Thoughtful test users should doubt the validity of IQ measures that identify this many children in an ordinary elementary school as candidates for special education.[4]

The expectancy effect disappears when extreme scores are omitted. The heightened experimental regression line (and thus the experimental mean) in the total group appears to result solely from five children whose respective pretest-posttest scores were 17-110, 18-122, 133-202, 111-208, and 113-211. If the small average difference in verbal subscores added anything to the total score difference, it was because of one experimental child whose pretest-posttest verbal scores were 133-202 (the same child whose reasoning scores were also 133-202).[5]

The second problem arises from failure to differentiate among studies of quite different variables in meta-analyses. Rosenthal summarized 464 expectancy studies with an effect size of .63, lumping together effects on Rorschach responses, reports of hallucinatory experiences, word association latencies, person perceptions, tone discriminations, verbal conditioning, and simple learning, along with the purported effects on IQ. But surely all these variables are not equally malleable. In particular, expectancy studies of ability should not be categorized with simple learning and conditioning studies to reach Rosenthal's reported effect size of .54 for the research domain he calls "learning and ability." Rather, the range of obtained effect sizes within and across categories suggests a continuum of proximal to distal variables that should be investigated: That is, an expectancy seems most likely to affect behavior that is in close proximity to it, and less likely to affect behavior that is more removed or distant from it, in time or space. Early reviews of teacher expectancy effects suggested such a continuum; most of the effects found were on teacher classroom behavior, some were on learner classroom behavior and achievement, but no effects on IQ appeared.[6] This result fits with voluminous other literature showing that mental abilities are not easily changed.

Each effect size entering the meta-analysis also needs to be studied carefully. Figure 2 shows Raudenbush's report of effect sizes for 18 studies of teacher expectancy effects on learner intelligence plotted as a function of estimated weeks of teacher-learner contact prior to expectancy induction.[7] Raudenbush noted the strong curvilinearity: Effect sizes are large only when prior contact between teacher and learner is 2 weeks or less. But three other points are also notable. First, there are 8 negative effect sizes (studies in which the control group gained more than the experimental group) against 10 positives. Second, Raudenbush reported a mean effect size of .11, but the distribution is highly skewed. Clearly, the median effect size of .035 is a more appropriate central tendency; without the discredited *Pygmalion* data, this median becomes .025. Third, at least one large effect size in Figure 2 appears questionable. Raudenbush entered .52 for the Pellegrini-Hicks study,[8] averaging .85 for a "tester aware" condition and .19 for a "tester blind" condition. However, the Pellegrini-Hicks study used testers who were all blind to the tests throughout, and compared individual tutors, not classroom teachers; the comparison was between tutors given both high expectations and familiarity with the tests being used to evaluate the tutoring (what Raudenbush called "tester aware"), tutors given high expectations but no test familiarity (Raudenbush's "tester blind"), and tutors given average or low expectations and no test familiarity. The differences due to expectations were minimal. Only test familiarization showed any effect, and only on one of two tests; the other test showed nothing, and high expectations alone showed nothing. Thus, the effect size of .85 probably comes mostly from test coaching. Raudenbush should have used the

effect size of .19 for high expectations alone, not the average of .52. And his comparison of aware versus blind testers cannot counter the possibility that *Pygmalion's* teacher-testers may have coached some students; consider again the five outlying asterisks in Figure 1!

A PROPOSED CURE

To cure the *Pygmalion* malady, I recommend following a threefold path. First, stop reporting *Pygmalion* as exemplary of diverse hundreds of studies, thereby misleading the public into thinking that intelligence can be changed just by changing expectations; tell the public it is not so. Second, distinguish and investigate the proximal-distal continuum of interpersonal expectancy effects instead of lumping everything together. Third, promote more careful thinking and detective work on the substantive psychology of the phenomenon and the methods and measures used to study it.

Notes

1. R. Rosenthal, Interpersonal expectancy effects: A 30-year perspective, *Current Directions in Psychological Science, 3,* 176–179 (1994); R. Rosenthal and D.B. Rubin, Interpersonal expectancy effects: The first 345 studies, *The Behavioral and Brain Sciences, 3,* 377–386 (1978). The original book by R. Rosenthal and L. Jacobson, *Pygmalion in the Classroom* (Holt, Rinehart and Winston, New York, 1968) has been reissued (Irvington, New York, 1992) with a cover banner indicating 50,000 copies have now been sold.

2. See, e.g., R.L. Thorndike, Review of *Pygmalion in the Classroom, American Educational Research Journal, 5,* 708–711 (1968), and Thorndike's exchange with R. Rosenthal in *American Educational Research Journal, 6,* 659–692 (1969); R.E. Snow, Unfinished Pygmalion, *Contemporary Psychology, 14,* 197–200 (1969); J.D. Elashoff and R.E. Snow, *Pygmalion Reconsidered* (Charles A. Jones, Worthington, OH, 1971), and the exchange with R. Rosenthal and D.B. Rubin included therein; J.B. Dusek, Ed., *Teacher Expectancies* (Erlbaum, Hillsdale, NJ, 1985); S.S. Wineburg, The self-fulfillment of the self-fulfilling prophesy, *Educational Researcher, 16*(9), 28–37 (1987), and the exchange with R. Rosenthal and R.C. Rist included therein.

3. Research shows substantial individual differences in susceptibility to expectancies for both teachers and learners. See, e.g., J. Brophy, Teacher-student interaction, in Dusek, note 2; J. Brophy and C. Evertson, *Student Characteristics and Teaching* (Longmans, New York, 1981).

4. See the exchange between Rosenthal and Thorndike, note 2. Rosenthal seemed not to understand Thorndike on this point.

5. Individual scores and scatterplots for all grades may be found in J.D. Elashoff and R.E. Snow, *A Case Study in Statistical Inference: Reconsideration of the Rosenthal-Jacobson Data on Teacher Expectancy,* Technical Report No. 15 (Stanford Center for Research and Development in Teaching, Stanford University, Stanford, CA, 1970).

6. See the review in Elashoff and Snow, note 2.

7. S.W. Raudenbush, 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 (1984).

8. R. Pellegrini and R. Hicks, Prophecy effects and tutorial instruction for the disadvantaged child, *American Educational Research Journal, 9,* 413–419 (1972).

Critiquing *Pygmalion:* A 25-Year Perspective

Robert Rosenthal

A quarter of a century ago, Richard E. Snow and his then collaborator, Janet D. Elashoff, devoted a book to trying to deny the Pygmalion effect.[1] Among their reanalyses of the original data, they tried eight variations, including some that were statistically biased.[2] Unfortunately for their position, every one of their reanalyses supported the original conclusions of the *Pygmalion* study. As Donald B. Rubin and I wrote in the last sentence of our reply, "The numerous criticisms advanced in ES [Elashoff and Snow] were neither sound nor constructive."[3] The same must be said of the present critique by Snow.

Robert Rosenthal is Edgar Pierce Professor of Psychology at Harvard University. Address correspondence to him at Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138.

On the basis of a literature review (that included no effect-size estimation), Snow concludes that no effects of teacher expectations on IQ have been demonstrated and that this fits the general finding that "mental abilities are not easily changed." If that statement is a tautology (i.e., "if the scores are easily changed, they are not mental ability scores"), nothing scientifically useful can be said in reply. If, however, this statement is meant to be a scientific statement about scores on measures of mental ability, it is wrong. Quite apart from Pygmalion effects, psychologists have known for half a century that such scores can be affected by the quality of social interaction in which the scores are obtained. In various studies, warmer examiners obtained IQs 6 points higher than did cooler examiners, and prior friendly contact led to an excess of 13 IQ points over scores in the control group. Similar results have been reported for other IQ-type measures, for signal detection, and for imaginativeness in Rorschach performance.[4]

In his analysis of the Raudenbush data, Snow points out that of all 18 studies, only 10 (56%) are positive in direction. This comment obfuscates the entire point of Raudenbush's brilliant moderator analysis of his Pygmalion meta-analysis. Among the 10 Pygmalion studies in