The Teacher as Pygmalion: Comments on the Psychology of Expectation

by Peter and Carol Gumpert

In Pygmalion in the Classroom, Robert Rosenthal and Lenore Jacobson present an experimental investigation of an interesting and provocative question: do teachers' expectations of how well, or badly, their students will perform actually influence the intellectual development of the students? The background of the study is introduced by a reminder that prophecies are sometimes instrumental in bringing about their own fulfillment. This reminder is followed by a discussion of evidence for the operation of self-fulfilling prophecies drawn from medical, psychiatric, and behavioral science research. The evidence compiled by Rosenthal and Jacobson is impressive: it is clear that what one person expects of another often has subtle, yet powerful, effects on the other. Since the teacher often "knows" (expects) which of her pupils will be brighter than the others, e.g., that white middle-class children will do better than the lower-class or Negro children, the critical question arises: can her expectation, by itself, bring about its own fulfillment?

To test the hypothesis that teacher expectations do, in fact, have such an effect, the investigators administered a group test of intelligence to several classes of children who were to enter grades one through six in 'Oak School.' The children's teachers were told that the test had been developed to predict forthcoming 'spurts' in intellectual growth; at the beginning of the fall term each teacher was given the names of several children in her classroom who could be expected to make unusual strides in intellectual development during the coming year. The specially designated children were actually no different from their contemporaries, having been selected randomly from the population of pupils at the school. The intelligence test was readministered to the children after a semester, after a year, and again after two years.

The major finding of the study was that the undesignated children, the control group, gained an average of more than eight IQ points after a year, while the special children, the experimental group, gained an average of more than 12. Statistical tests suggested that this difference was one that could not reasonably be ascribed to chance variation alone. Rosenthal and Jacobson concluded—and we concur—that their major hypothesis was confirmed.

They also report a great deal of additional data on the more specific effects of the expectations they induced among the teachers of Oak School. Some of these ancillary results are intriguing, particularly some of the observations concerned with teachers' perceptions of their students during the first year of the experiment. However, looking carefully at all the results reported by Rosenthal and Jacobson—how the strength of teacher expectation varied with the sex of the students, their group status (minority or otherwise), their ability track, and so on—is a complex and confusing business. It is fairly easy to interpret each result taken alone, but the reader who attempts to make his interpretations consistent with one another will be in for a good deal of difficulty. The situation becomes even more confusing when we add to the Oak School results those obtained in the three similar studies in other schools reported by Rosenthal and Jacobson (pp. 96, 138, 145). Here we find apparent reversals of some of the effects that are given much emphasis in the Oak School study. One of these, to which Rosenthal and Jacobson pay special attention, is the result that after one year, Oak School girls in the experimental group showed greater gains on the reasoning subtest of the IQ test than did boys, while the reverse was the case on the verbal subtest, where boys showed evidence of greater improvement than girls. Since initially boys had shown slightly higher verbal scores and girls had shown slightly higher reasoning scores, Rosenthal and Jacobson interpret their results as indicating that each group tends to benefit most from favorable teacher expectations in those areas in which it is somewhat advantaged at the outset. However, in a previous experiment, while the initial scores for the sexes show the same pattern, the pattern of gains is reversed.

Rosenthal and Jacobson respond to this by postulating that their explanation for the Oak School result is not applicable to the earlier experiment, presumably because the schools in that study drew their pupils from a 'substantial middle-class community,' while Oak School's pupils came from a lower-class community. They do not explain how the socioeconomic class difference could account for the discrepancy in results between the two studies. The authors do not abandon their faith in the relationship between sex and subtest in either of the two studies, even though the results are by most people's standards mutually contradictory. They see the discrepancy in results as another demonstration of the "probable complexity of the operation of teacher expectancy effects" (p. 162).

This points to what we consider to be a major flaw of the book. The authors somehow give the impression that the 'expectancy effect' is a mysterious and profoundly complex entity, defying attempts by behavioral science to account for it. From such a viewpoint, it is not surprising that the importance of statistical significance is overemphasized by these investigators while the importance of replicability is underestimated. They see the discrepancy in results as another demonstration of the complexity of the operation of teacher expectancy effects" (p. 162).

Some of the confusion disappears if one takes another approach to Rosenthal and Jacobson's statistical analysis. In the Oak School experiment each child in each of 18 classrooms was used either as an "experimental" or a "control" subject in the study. Clearly, this was the easiest and most natural way for the researchers to obtain a reasonably large sample of children. But children affect one another, and many things teachers and children do can affect classroom atmosphere; so there is reason to suspect that each child's test gains cannot be seen as an independent observation for statistical purposes. If the assumption of independence were, in fact, not tenable, we could conclude that they overestimated the degree of statistical stability of much of their data. Accordingly, we reanalyzed some of their major results using the analysis of variance to dis-
cover whether a child’s performance gain could be partially accounted for by the particular classroom he happened to be in, independent of his grade level or whether he was in the experimental or the control group. As we suspected, the effect of classroom membership is very stable in the analyses we did (p < .001), which means that strictly speaking, the proper “unit of analysis” for the study is the classroom average rather than the individual child (see the footnotes on pp. 95, 96, 118, 119, 145, and 146). Looking at the data this way, many of the results that were statistically significant using the analysis of variance as Rosenthal and Jacobson used it do not reach significance when the more usual and, in fact, more proper test is applied.

On the other hand, even if we confine Rosenthal and Jacobson to these more conservative statistical techniques, we find that the most important results of their study remain statistically stable. Further, these results are reasonably easy to interpret consistently. First, it remains clear that a semester after the teachers had been given information about the “sputters” in their classrooms, the children designated as sputters showed somewhat greater improvement on the intelligence test than did the control children. This overall effect of teachers’ expectancies became much greater at the stable) at the end of two years. By and large, these effects were more pronounced in the first and second grades than they were in the later grades.

Second, the experimental children tended to show greater gains in reading grades on their report cards and were judged by their teacher to be, on the average, more intellectually curious than the control children. Finally, the average degree of improvement shown by the experimental children in a classroom was strongly associated with the average degree of improvement shown by the control children in the same classroom: if the experimental children did especially well, so did the control children. So it would seem that the general classroom ‘atmosphere’ generated by all concerned has something to do with what happens. The second and third set of results are of particular interest because they provide us with clues as to how some of the classroom expectancy effects observed in the Oak School study might have taken place, a consideration to which we shall return.

Let us note what we consider to be some of the book’s other weaknesses. First, Rosenthal and Jacobson speak repeatedly of “IQ gains,” thus implying that changes occurred in children’s intellectual “capacity” or intellectual functioning as this is usually understood by the layman. Such a claim would, of course, be unjustified. A discussion of the nature of intelligence and theories about its measurement is not possible within the scope of this paper. Suffice it to say that a change in IQ score can be interpreted simply as a change in performance on a test for which age-norms exist. Since rather little normative information is available (especially on younger children) for the intelligence test used in their research, the IQ gains reported by Rosenthal and Jacobson must be interpreted with particular caution.

Second, there is implicit confusion throughout the book about the difference between the “strength” of a variable and the level of statistical significance associated with its effects. Though the two things often go together, it is important to note that a statistical significance level is an index of the probable stability of a result, and not of its magnitude. A very weak effect that is extremely stable will be detected at a high level of statistical significance even though it may not be very meaningful in practical terms. If, for example, we were to weigh a large number of automobiles on a very sensitive scale, add a thimbleful of gasoline to the gas tanks of most of them, and weigh them again, the changes in their weight would turn out to be highly significant statistically, though the thimblefuls of gasoline would still have contributed little, relatively speaking, to the total weight of the automobiles. In assessing the importance of a variable, then, one must take into account not only the stability and the apparent magnitude of its effects, but also the significance of a result of that magnitude in the total context of the research.

These points aside, Pygmalion in the Classroom is an important and thought-provoking book; anyone concerned with the problems and practices of education should certainly take the trouble to read it. The study provides a perfectly satisfactory demonstration that the teacher expectancy effects hypothesized do indeed take place. Though the Oak School experiment is not as sophisticated as it might be, it was done in a ‘natural’ setting (which, incidentally, usually makes it difficult to do elegant research) rather than under the more artificial circumstances of the laboratory. The fact that the effect can be demonstrated in the actual classroom under very ordinary conditions is convincing. That the effect appears to be quite strong and stable in spite of the subtlety and simplicity of the experimental induction is especially dramatic. The results of the experiment fairly demand that much serious attention, thought, and research be devoted to the effect on children of the beliefs and attitudes held about them by school administrators, supervisors, and teachers. It also points up the crucial importance of conducting research on just how people’s expectations of children will become realities. It is possible that we will learn to change teachers’ behavior toward children before we learn how to change their attitudes toward them, and thus their expectations of them. It is on this last general consideration—how the expectation of the teacher might have led to modifications of her pupils’ performance and classroom behavior—that Rosenthal and Jacobson are weakest in their analysis. Though they do speculate about some aspects of the problem, the heart of the matter remains untouched. In short, they have shown us that teachers’ expectations of their students’ performances have definite consequences for these (subsequent) performances. But they have not shown us how this process works.

In the remainder of our article, we propose to do some more or less systematic speculating about how the Oak School teachers may have fulfilled the researchers’ prophecies. We shall begin by arguing that expectation leads to selectivity of attention, perception, and response, and end by discussing just how increases in interpersonal “warmth” and encouragement might actually lead to superior learning and performance.

**Some Psychological Effects of Expectation**

The study of the influence of expectation upon thinking and behavior has been of interest to psychologists for many years in a variety of contexts. There are literally hundreds of studies that are relevant to the notion that a person’s attitude, set, or expectation will affect his perceptions and responses. One tentative conclusion that can be drawn from these studies—of particular interest in thinking about expectancy in the classroom—is the following: a person is more likely to perceive a barely perceptible stimulus if he expects it than if he does not. For example, if an experimental subject is given a list from which words are to be presented tachistoscopically, the recognition threshold for words on the list (i.e., the words he is expecting) will be lower than for those not on the list (Neisser, 1954). A related study is reported by Bruner and Minturn (1955). When shown an ambiguous stimulus which resembled both a letter and a number, for example B and