36

Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development

Robert Rosenthal and Lenore Jacobson

There is increasing concern over what can be done to reduce the disparities of education, of intellectual motivation, and of intellectual competence that exist between the social classes and the colors of our school children. With this increasing concern, attention has focused more and more on the role of the classroom teacher, and the possible effects of her or his values, attitudes, and, especially, beliefs and expectations. Many educational theorists have expressed the opinion that the teacher's expectation of her pupils' performance may serve as an educational self-fulfilling prophecy. The teacher gets less because she expects less.

The concept of the self-fulfilling prophecy is an old idea which has found application in clinical psychology, social psychology, sociology, economics, and in everyday life. Most of the evidence for the operation of self-fulfilling prophecies has been correlational. Interpersonal prophecies have been found to agree with the behavior that was prophesied. From this, however, it cannot be said that the prophecy was the cause of its own fulfillment. The accurate prophecy may have been based on a knowledge of the prior behavior of the person whose behavior was prophesied, so that the prophecy was in a sense "contaminated" by reality. If a physician predicts a patient's improvement, we cannot say whether the doctor is only giving a sophisticated prognosis or whether the patient's improvement is based in part on the optimism engendered by the physician's prediction. If school children who perform poorly are those expected by their teachers to perform poorly, we cannot say whether the teacher's expectation was the "cause" of the pupils' poor performance, or whether the teacher's expectation was simply an accurate prognosis of performance based on her knowledge of past performance. To help answer the question raised, experiments are required in which the expectation is experimentally varied and is uncontaminated by the past behavior of the person whose performance is predicted.

Such experiments have been conducted and they have shown that in behavioral research the experimenter's hypothesis may serve as self-fulfilling prophecy (Rosenthal, 1966). Of special relevance to our topic are those experiments involving allegedly bright and allegedly dull animal subjects. Half the experimenters were led to believe that their rat subjects had been specially bred for excellence of learning ability. The remaining experimenters were led to believe that their rat subjects were genetically inferior. Actually,

"Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development" by R. Rosenthal and L. Jacobson from *Doing Unto Others* edited by Zick Rubin (pp. 41–47), 1974. Englewood Cliffs, NJ: Prentice-Hall. Reprinted by permission of the authors.

the animals were assigned to their experimenters at random.

Regardless of whether the rat's task was to learn a maze or the appropriate responses in a Skinner box, the results were the same. Rats who were believed by their experimenters to be brighter showed learning which was significantly superior to the learning by rats whose experimenters believed them to be dull. Our best guess, supported by the experimenters' self-reports, is that allegedly well-endowed animals were handled more and handled more gently than the allegedly inferior animals. Such handling differences, along with differences in rapidity of reinforcement in the Skinner box situation, are probably sufficient to account for the differences in learning ability shown by allegedly bright and allegedly dull rats.

If rats showed superior performance when their trainer expected it, then it seemed reasonable to think that children might show superior performance when their teacher expected it. That was the reason for conducting the Oak School Experiment.

The Oak School Experiment

To all of the children in the Oak School, on the West Coast, the "Harvard Test of Inflected Acquisition" was administered in the Spring of 1964. This test was purported to predict academic "blooming" or intellectual growth. The reason for administering the test in the particular school was ostensibly to perform a final check of the validity of the test, a validity which was presented as already well-established. Actually, the "Harvard Test of Inflected Acquisition" was a standardized, relatively nonverbal test of intelligence, Flanagan's Tests of General Ability.

Within each of the six grades of the elementary school, there were three classrooms, one each for children performing at above-average, average, and below-average levels of scholastic achievement. In each of the 18 classrooms of the school, about 20% of the children were designated as academic "spurters." The names of these children were reported to their new teachers in the Fall of 1964 as those who, during the academic year ahead, would show unusual intellectual gains. The "fact" of their intellectual potential was established from their scores on the test for "intellectual blooming."

Teachers were cautioned not to discuss the test findings with either their pupils or the children's parents. Actually, the names of the 20% of the children assigned to the "blooming" condition had been selected by means of a table of random numbers. The difference, then, between these children, earmarked for intellectual growth, and the undesignated control group children was in the mind of the teacher.

Four months after the teachers had been given the names of the "special" children, all the children once again took the same form of the nonverbal test of intelligence. Four months after this retest the children took the same test once again. This final retest was at the end of the school year, some eight months after the teachers had been given the expectation for intellectual growth of the special children. These retests were not explained as "retests" to the teachers, but rather as further efforts to predict intellectual growth.

The intelligence test employed, while relatively nonverbal in the sense of requiring no speaking, reading, or writing, was not entirely nonverbal. Actually there were two subtests, one requiring a greater comprehension of English—a kind of picture vocabulary test. The other subtest required less ability to understand any spoken language but more ability to reason abstractly. For shorthand purposes we refer to the former as a "verbal" subtest and to the latter as a "reasoning" subtest. The pretest correlation between these subjects was only +.42, suggesting that the two subtests were measuring somewhat different intellectual abilities.

For the school as a whole, the children of the experimental groups did not show a significantly greater gain in verbal IQ (2 points)

than did the control group children. However, in total IQ (4 points) and especially in reasoning IQ (7 points) the experimental children gained more than did the control group children. In 15 of the 17 classrooms in which the reasoning IQ posttest was administered, children of the experimental group gained more than did the control group children. Even after the four-month retest this trend was already in evidence though the effects were smaller.

When we examine the results separately for the six grades we find that it was only in the first and second grades that children gained significantly more in IQ when their teacher expected it of them. In the first grade, children who were expected to gain more IQ gained over 15 points more than did the control group children. In the second grade, children who were expected to gain more IQ gained nearly 10 points more than did the control group children. In the first and second grades combined, 19% of the control group children gained 20 or more IQ points. Two-and-a-half times that many, or 47%, of the experimental group children gained 20 or more IQ points.

When educational theorists have discussed the possible effects of teachers' expectations, they have usually referred to the children at lower levels of scholastic achievement. It was interesting, therefore, to find that in the present study, children of the highest level of achievement showed as great a benefit as did the children of the lowest level of achievement of having their teachers expect intellectual gains.

At the end of the school year of this study, all teachers were asked to describe the classroom behavior of their pupils. Those children from whom intellectual growth was expected were described as having a significantly better chance of becoming successful in the future, as significantly more interesting, curious, and happy. There was a tendency, too, for these children to be seen as more appealing, adjusted, and affectionate and as lower in the need for social approval. In short, the children from whom intellectual growth was expected became more intellectually alive and autonomous—or at least were so perceived by their teachers. These findings were particularly striking among first-grade children; these were the children who had benefited most in IQ gain as a result of their teachers' favorable expectancies.

We have already seen that the children of the experimental group gained more intellectually. It was possible, therefore, that their actual intellectual growth accounted for the teachers' more favorable ratings of these children's behavior and aptitude. But a great many of the control group children also gained in IQ during the course of the year. Perhaps those who gained more intellectually among these undesignated children would also be rated more favorably by their teachers. Such was not the case. In fact, there was a tendency for teachers to rate those control group children who gained most in IQ as *less* well-adjusted, *less* interesting, and *less* affectionate than control group children who made smaller intellectual gains. From these results it would seem that when children who are expected to grow intellectually do so, they may benefit in other ways as well. When children who are not especially expected to develop intellectually do so, they may show accompanying undesirable behavior, or at least are perceived by their teachers as showing such undesirable behavior. It appears that there may be hazards to unpredicted intellectual growth.

A closer analysis of these data, broken down by whether the children were in the high, medium, or low ability tracks or groups, showed that these hazards of unpredicted intellectual growth were due primarily to the children of the low ability group. When these slow track children were in the control group, so that no intellectual gains were expected of them, they were rated less favorably by their teachers if they did show gains in IQ. The greater their IQ gains, the less favorably were they rated, both as to mental health and as to intellectual vitality.

Even when the slow track children were in the experimental group, so that IQ gains were expected of them, they were not rated as favorably relative to their control group peers as were children of the high or medium track, despite the fact that they gained as much in IQ relative to the control group children as did the experimental group children of the high track. It may be difficult for a slow track child, even one whose IQ is rising, to be seen by his teacher as a well-adjusted child, or as a potentially successful child intellectually.

The Question of Mediation

How did the teachers' expectations come to serve as determinants of gains in intellectual performance? The most plausible hypothesis seemed to be that children for whom unusual intellectual growth had been predicted would be attended to more by their teachers. If teachers were more attentive to the children earmarked for growth, we might expect that teachers were robbing Peter to see Paul grow. With a finite amount of time to spend with each child, if a teacher gave more time to the children of the experimental group, she would have less time to spend with the children of the control group. If the teacher's spending more time with a child led to greater intellectual gains, we could test the "robbing Peter" hypothesis by comparing the gains made by children of the experimental group with gains made by the children of the control group in each class. The robbing Peter hypothesis predicts a negative correlation. The greater the gains made by children of the experimental group (with the implication of more time spent on them) the less should be the gains made by the children of the control group (with the implication of less time spent on them). In fact, however, the correlation was positive, large, and statistically significant (+.57). The greater the gains made by children of whom gain was expected, the greater the gains made in the same classroom by those children from whom no special gain was expected.

Additional evidence that teachers did not take time from control group children to spend with the experimental group children comes from the teachers' estimates of time spent with each pupil. These estimates showed a slight tendency for teachers to spend *less* time with pupils from whom intellectual gains were expected.

That the children of the experimental group were not favored with a greater investment of time seems less surprising in view of the pattern of their greater intellectual gains. If, for example, teachers had talked to them more, we might have expected greater gains in verbal IQ. But the greater gains were found not in verbal but in reasoning IQ. It may be, of course, that the teachers were inaccurate in their estimates of time spent with each of their pupils. Possibly direct observation of the teacher-pupil interactions would have given different results, but that method was not possible in the present study. But even direct observation might not have revealed a difference in the amounts of teacher time invested in each of the two groups of children. It seems plausible to think that it was not a difference in amount of time spent with the children of the two groups which led to the differences in their rates of intellectual development. It may have been more a matter of the type of interaction which took place between the teachers and their pupils.

By what she said, by how she said it, by her facial expressions, postures, and perhaps by her touch, the teacher may have communicated to the children of the experimental group that she expected improved intellectual performance. Such communications, together with possible changes in teaching techniques, may have helped the child learn by changing his or her self-concept, expectations of his or her own behavior, motivation, as well as cognitive skills. Further research is clearly needed to narrow down the range of possible mechanisms whereby a teacher's ex-

pectations become translated into a pupil's intellectual growth. It would be valuable, for example, to have sound films of teachers interacting with their pupils. We might then look for differences in the ways teachers interact with those children from whom they expect more intellectual growth compared to those from whom they expect less. On the basis of films of psychological experimenters interacting with subjects from whom different responses were expected, we know that even in such highly standardized situations, unintentional communications can be subtle and complex (Rosenthal, 1967). How much more subtle and complex may be the communications between children and their teachers in the less highly standardized classroom situation?

Conclusions

The results of the Oak School experiment provide further evidence that one person's expectations of another's behavior may serve as a self-fulfilling prophecy. When teachers expected that certain children would show greater intellectual development, those children did show greater intellectual development. A number of more recent experiments have provided additional evidence for the operation of teacher expectancy effects, in contexts ranging from the classroom to teaching athletic skills. Although not all of the studies that have been conducted show such effects, a large proportion of them do (Rosenthal, 1971).

It may be that as teacher training institutions acquaint teachers-to-be with the possibility that their expectations of their pupils' performance may serve as self-fulfilling prophecies, these teacher trainees may be given a new expectancy—that children can learn more than they had believed possible.

Perhaps the most suitable summary of the hypothesis discussed in this paper has already been written. The writer is George Bernard Shaw, the play is *Pygmalion,* and the speaker is Eliza Doolittle:

> You see, really and truly, . . . the difference between a lady and a flower girl is not how she behaves, but how she's treated. I shall always be a flower girl to Professor Higgins, because he . . . treats me as a flower girl, . . . but I know I can be a lady to you, because you always treat me as a lady, and always will.

Note

An expanded discussion of self-fulfilling prophecies and a full account of the Oak School experiment are presented in R. Rosenthal & L. Jacobson (1968), *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development,* New York: Holt, Rinehart & Winston.

References

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R. (1967). Covert communication in the psychological experiment. *Psychological Bulletin, 67,* 356–367.

Rosenthal, R. (1971). Teacher expectation and pupil learning. In R. D. Strom (Ed.), *Teachers and the learning process.* Englewood Cliffs, NJ: Prentice-Hall.