DOCUMENT RESUME

ED 080 567 TM 003 081

AUTHOR Wilkins, William E.; Glock, Marvin D.

TITLE Teacher Expectations and Student Achievement: A Replication and Extension. Final Report.

INSTITUTION Cornell Univ., Ithaca, N.Y.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO BR-1-B-050
PUB DATE Jun 73
GRANT OEG-2-71-0050
NOTE 65p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Gains; Achievement Tests; Aptitude Tests; Elementary Grades; *Expectation; Literature Reviews; *Low Achievement Factors; Student Teacher Relationship; *Teacher Behavior; *Teacher Influence; Technical Reports

ABSTRACT

Critiques of the Rosenthal research on experimenter bias and teacher expectancy have demonstrated that few, if any, conclusions can be drawn from it. From other research concerning the factors in and effects of teacher expectancy, it appears that elementary school children make accurate perceptions of subtle affective and cognitive behaviors of teachers and peers. Furthermore, it seems reasonable that pupils can perceive differential expectations of the teacher and that these perceptions may be a vital link in whatever communication system may be involved in the self-fulfilling prophecy process. This study utilized all students in 24 classrooms, grades 1-6, who were administered achievement and learning potential tests at the beginning of the school year. After teachers ranked their students according to how much achievement growth they expected from them, experimental pairs were selected and members of the pairs were randomly assigned to treatment and control groups. A questionnaire to discern pupils' perceptions of teachers' differential expectations and treatment of students was also administered. Teachers were told that test scores indicated that certain students should advance much more rapidly than the teachers had anticipated. Observations were made in each classroom. No significant effects of experimental manipulation of teacher expectations were found. It was concluded, therefore, that the Rosenthal findings are spurious. (For related document, see TM 003 083.) (KM)

FINAL REPORT
Project No. 1-B-060
Grant No. OEG-2-71-0060

TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT: A REPLICATION AND EXTENSION

William E. Wilkins
Marvin D. Glock
Cornell University
Ithaca, New York

June, 1973

U.S. DEPARTMENT OF
HEALTH, EDUCATION AND WELFARE

Office of Education
Bureau of Research

Final Report
Project No. 1-B-050
Grant No. OEG-2-71-0050

Teacher Expectations and Student Achievement: A Replication and Extension

William E. Wilkins
(Principal Investigator)
Department of Psychology
SUC at Brockport
Brockport, New York

Marvin D. Glock
(Project Director)
Department of Education
Cornell University
Ithaca, New York

June, 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

CHAPTER I

THE INFLUENCE OF TEACHER EXPECTATION ON PUPIL ACHIEVEMENT

A. Experimenter Bias:

The current interest in experimenting with the effects of teacher expectations evolved out of the work on experimenter bias by Robert Rosenthal. Rosenthal (1966) documented the case for observer expectations or experimenter bias in the physical sciences, the biological sciences, and the behavioral sciences.

During a decade of experimental investigation, Rosenthal and his co-workers have shown that exoerimenter expectations have a profound effect even under rigorous laboratory circumstances. This effect, which Rosenthal terms a self-fulfilling prophecy, has been shown to influence results when animals were used as subjects (eg. Rosenthal and Fode, 1963; and Burhaus, 1966). Similar findings have been reported and summarized by Rosenthal (1966, 1968) using human subjects. Illustrative of the animal experiments is a study conducted by Burnham (1966). In this study, the investigator had twenty-three experimenters condition rats in a simple T-maze discrimination task. Approximately one-half of the rats were brain-lesioned by surgical removal of the cortex. The remaining animals underwent surgery but the brain was left untouched. Expectancies were manipulated by telling the experimenters that they were testing the effects of brain lesion on discrimination learning. The entire sample of rats was then arbitrarily labeled lesioned or non-lesioned with some

lesioned being labeled non-lesioned and vise-virsa. Results supported the hypotheses. Those rats not lesioned and labeled non-lesioned performed the best, all rats which were actually lesioned performed worst irregardless of expectation. Most importantly, however, those rats which were actually not lesioned but were labeled as lesioned, performed as poorly as the in-fact lesioned rats. Thus, rats expected to perform well did, while those expected to perform poorly did in fact perform poorly.

A variety of other studies using rats (eg. Rosenthal and Fode, 1963; Ingraham and Harrington, 1966) and planaria (Cordaro and Ison, 1963) support Rosenthal's hypotheses under laboratory situations. While Rosenthal and his co-workers have been unable to specifically determine the precise mechanisms of experimenter bias with animals, indications are that those experimenters who believed that their animals were going to do well spent more time handling their animals, whereas those who expected their animals to do poorly, spent more time talking to their animals.

Representative of research on experimenter bias using human subjects is a study by Rosenthal and Fode (1961). In this investigation, ten experimenters had 206 subjects judge the probable success or failure of people in ten photographs on a ten point scale (from +5 to -5) from extreme failure to extreme success. Half of the experimenters were led to believe that most people rated the ten photos on the positive side, whereas the remaining five were told that people tended to judge the photos on the failure end of the continuum. Previous standardization of the photos indicated an average rating of 0. As hypothesized,

the ratings of subjects contacted by the experimenter expecting negative ratings tended to get negative ratings, and conversely, experimenters expecting positive ratings tended to get positive ratings ($t=3.20$, $df=8$ $p>.007$ one tailed). Confirmation of experimenter expectancies with humans has been attributed to some, as yet unknown, subtle communication systems between experimenter and subject. Although Rosenthal and his colleagues have discerned certain physiological and demographical chatacteristics (eg. sex, age, race, and nationality) which correlate with the self-fulfilling prophecy, enumerable man hours of observing experiments, live and video taped, have not yielded a detailed specification of the elements of this intercourse.

These findings, however, have been criticized by Barber and Silver (1968), Barber (1969), Barber, et.al. (1969), and Barber and Silver (1969). Although these investigators believe that experimenter bias may exist, they emphasize that their own research (eg. Barber, et al., 1969) does not support Rosenthal's conclusions. Furthermore, Barber and his co-workers contend that the methods of data analysis utilized by the Rosenthal group are suspect. In fact, argue Barber and Silver (1969), out of thirty-one studies reported on experimenter bias, only twelve support Rosenthal's thesis; and of these twelve, only three provide specific evidence for experimenter bias. While Barber agrees that experimenter bias may exist, he contends that it is a difficult variable to demonstrate and that Rosenthal's apparent clear-cut data may be the result of improper analysis.

Nonetheless, Rosenthal proceeded to generalize his experimenter

bias hypothesis to the teaching-learning situation. In developing a rationale for their teacher expectation study, Rosenthal and Jacobson (1968) write:

> The transition from this chapter to the next (animal studies to teacher expectation studies) comes from the results of the experiment in which rats learned their way around their Skinner boxes. At the beginning of that study experimenters assigned allegedly dull animals were of course told that they would find retarded learning on the part of their rats. They were, however, reassured that "it has been found that even the dullest rats can, in time, learn the required responses" (Rosenthal and Lawson, 1964). Animals alleged to be dull, then were described as educable but slow. It was interesting in the light of this to learn that of the experimenters who had been assigned "dull" animals, 47 percent believed their subjects to be uneducable. Only 5 percent of the experimenters assigned "bright" rats were equally pessimistic about their animal's future (p=.007). From this result one wonders about the beliefs created in school-teachers when they are told a child is educable but slow, deserving but disadvantaged (p. 44).

B. The Oak School Experiment:

Of specific relevance for the present study is Rosenthal's and Jacobson's application of the self-fulfilling prophecy concept to the elementary education process (Rosenthal and Jacobson, 1966). Their study, known as the Oak School Experiment, was later duplicated and expanded as part of a manuscript intitled Pygmalion in the Classroom (Rosenthal and Jacobson, 1968). In this experiment, Rosenthal and Jacobson administered Flanagan's Tests of General Ability to all pupils in grades K-6 in the Spring of 1964. Classroom teachers were told that this test was the "Harvard Test of Inflected Acquisition" and that further validation was necessary. Eighteen

classrooms, three at each of the six grade levels, were selected to participate in the experiment. Each grade level was represented by a low ability group, an average ability group, and a high ability group. A random sample of 20 percent of the pupils in each of the 18 classrooms was designated as potential academic "spurters." That is, when the pupils returned to school the following year, classroom teachers were told that certain of their pupils, designated "bloomers," would make unusual academic gains. There was retesting at the middle of the academic year and post testing at the end of the school year. Final retesting was completed during the following Spring.

Rosenthal and Jacobson (1968) report inconsistent findings. At one point they argue that significant gains were found at all six grade levels (p. 74), and at other times they report significant gains in grades one and two only (p. 77). A review of the data presented by Rosenthal and Jacobson shows, in fact, significant gains for 19 children in grades one and two. This fact led Thorndike (1968), among many others, to be sharply critical of the Oak School findings. Noting that "to all intents and purposes, the alleged effect of the 'prophecy' appears in 19 children in grades one and two," Thorndike insists that "if we are to trust the results, and the large edifice of further analysis and speculation built upon them, the findings of these two grades must be unimpeachable" (p. 709). Thorndike then proceeds to analyze the IQ measures and scores reported by Rosenthal and Jacobson and concludes that "if the pretest data show anything, they show that the

testing was utterly worthless and meaningless for grades one and two." Thorndike ends his review by stating: "the indications are that the basic data upon which this structure has been raised are so untrustworthy that any conclusions based upon them must be suspect. The conclusions may be correct, but if so, it must be considered a fortunate coincidence (p. 711)."

An extensive and equally critical review of Pygmalion in the Classroom is presented by Elashoff and Snow (1971). In this critique, Elashoff and Snow discuss several faults including data analysis, data presentation, the testing instruments, and the research methodology. Noting that "the research report is a crucial part of the research process," Elashoff and Snow contend that Pygmalion in the Classroom is an inadequate report because "descriptions of design, basic data, and analysis are incomplete. Inconsistencies between text and tables, overly dramatic conclusions, oversimplified, inaccurate or incorrect statistical discussions and analyses all contribute to a generally misleading impression of the study's results" (p. 6).

Specifically, the objections concern the use of, and interpretation of, the Flanders Test of General Abilities (TOGA). Agreeing with Thorndike, Elashoff and Snow note that the supposed effect occured only in grades one and two. Furthermore, it becomes apparent upon examining the scores of the first and second graders, that many did not seem to understand the operations expected of them. That is, several of the Rosenthal and Jacobson experimental subjects had pretest IQ scores which

were obviously incorrect and uninterpretable. One child had a pretest IQ of 17, another a score of 30, and some experimental subjects had post test IQs of 212 and 220. Elashoff and Snow argue that if we eliminate those with pretest IQs around 30 and those with post test IQs of 220 and 230, no expectancy effect occurs. They point out that IQ gains to 90 points, from 30 to 120 or from 140 to 230, are impossible to interpret and therefore cannot be meaningfully included in data analysis.

It should be noted however, that it is a difficult and questionable practice to decide which scores to eliminate from any data analysis. To be sure, it seems obvious that IQ scores of 30 for children admitted to first grade are unreasonable; yet if we begin to arbitrarily truncate our data, what guidelines can we use?

Other specific criticisms of the Oak School Experiment include questions of experimental design and methodology. Elashoff and Snow argue that, because of subject loss and assignment to experimental condition based on the entire school population rather than grade level or ability level, it is questionable whether the Rosenthal and Jacobson subjects were randomly selected. Closely allied with this criticism is another which questions the lack of control subjects. Elashoff and Snow maintain that some semblence of a randomized block design which took such factors as age, sex, grade and ability into account would have been more appropriate and meaningful. Furthermore, Rosenthal and Jacobson's selection of an ex post facto control group is of dubious value.

One further issue which suggests spuriously derived results is the fact that, after the experiment was over, the teachers could neither recall or recognize the names of the "bloomers" that Rosenthal and Jacobson had given them at the onset of the experiment. Since the teachers could not identify those students they were supposedly giving "special" treatment, it seems difficult to account for IQ gains on the basis of the experimentally "induced" expectations.

The conclusions reached by Rosenthal and Jacobson are also suspect from the work of other investigators. Numerable replications and near replications have failed to support the findings reported in Pygmalion in the Classroom. Further, previous attempts to demonstrate teacher expectancy also lend opposing evidence to the Rosenthal and Jacobson findings. Pitt (1956) and Flowers (1966), failed to find any significant relationships between teacher expectancy and a variety of achievement and socio-psychological factors. These studies differed from the Oak School Experiment (eg. Pitt reassigned entire classrooms of low ability pupils as high ability pupils), and are not directly comparable; therefore, while the findings of Pitt and Flowers cast some doubt on the Rosenthal and Jacobson findings, they cannot be considered conclusive. Furthermore, the subjects in both of these studies were older children (grade 7), and Rosenthal and Jacobson did not find significant gains among pupils in grades 3-6, although significant gains were reported for grades 1 and 2. While Rosenthal interprets the grade level difference in the Pitt and Flowers studies as

not contradictory to his findings, one should recall that he is not consistent in his grade level interpretations.

The findings thus far discussed i . that the data reported by Rosenthal and Jacobson may well be specious. Data reported by other investigators helps to clarify the Rosenthal and Jacobson conclusions and the subsequent criticisms of those conclusions. One fact that has emerged is the necessity of differentiating between gains in achievement and gains in IQ. That is, while some investigators have found significant gains in achievement test scores, no other investigator has been able to increase IQ scores as a function of experimental manipulation of teacher expectations.

The following studies have been reported using IQ measures .s the sole or one dependent variable. The independent variable in all of these studies was some form of manipulation of teacher expectation. Conn, Edwards, Rosenthal and Crowne (1968) tried to replicate the Oak School findings and to study the underlying processes involved. Using the TOGA and the usual Rosenthal expectancy induction, Conn, et al., found no significant IQ increases. However, they did find a significant relationship between expectancy gains and perception of emotion in the teacher's voice, especially for boys in the experimental group.

Evans and Rosenthal (1969), using grades 1 through 6, attempted another partial replication of the Rosenthal and Jacobson study. Again, using the TOGA and the typical expectancy induction, no significant main effects were found. One interaction effect was found with Verbal IQ.

Claiborn (1969) attempted a replication using 12 first grade classrooms. With IQ gains on the TOGA and classroom observations serving as the dependent variables, Claiborn found no significant effects for any variables. It should be noted that Claiborn and Conn, et al., tried to induce expectancy during the middle of the school year.

Other attempts to replicate the Rosenthal and Jacobson experiment have been equally unsuccessful. José and Cody (1971), using 18 first and second grade classrooms and a randomized block design, found no differences between experimental and control subjects on IQ scores (TOGA), achievement test scores (Metropolitan Achievement Test), or pupil-teacher interactions. José suggested, among other things, that simply telling the teacher that certain students were going to bloom and showing them IQ scores may not be sufficient to induce expectancy.

Fleming and Anttonen (1971) used achievement scores, self concept scores, IQ, and teacher grading behavior as criterion measures in their investigation of the effects of differing teacher expectations on a sample of 900 second grade students. Four conditions were established on the basis of information given to the classroom teachers: 1) IQ scores inflated 16 points (Kuhlmann-Anderson IQ Test); 2) actual Kuhlmann-Anderson IQ scores; 3) actual scores on the Primary Mental Abilities Test (PMA); and 4) no information. Children were randomly assigned to one of the four conditions within each classroom. No significant expectancy effects in either measures of IQ or measures of self concept were found. Inter-

action effects between SES of the children and the teacher's opinion of tests did lead to some generally inconsistent significant effects. For example, teachers who held higher regard for tests gave higher grades and some expectancy effects as revealed by somewhat higher IQ gains and achievement test gains when compared with teachers who held tests in a generally low regard.

In a study using older children (grade 7), Kester (1971) randomly assigned 150 average ability students to experimental and control conditions. The twenty three teachers participating in the study were given the names of the "bright" students in their class during the first week of school. Criterion measures in this study were the Stanford Achievement Test (SAT), the Otis-Lennon IQ Test, and a measure of pupil attitudes. Observations were made under the ruse of watching teacher-"bright" student interactions. No significant differences were found between pretest and post test scores on the dependent variables; however, teachers did seem to interact more, and in a more positive fashion, with the experimental Ss.

Utilizing a matched sample of 112 experimental and 112 control Ss, Goldsmith and Fry (1970) attempted to demonstrate expectancy effects with high school students. Dependent variables were IQ scores (TOGA) and scores on the Sequential Test of Educational Progress (STEP-2B). To counteract any forgetting effects, the teachers were reminded of the names of the "bloomers" from three to five times during the course of the experiment. Analysis of gains revealed no significant effects

on either the TOGA or the STEP.

Two attempts to demonstrate expectancy effects on IQ outside the regular classroom situation have been reported. Anderson and Rosenthal (1968) used 28 institutionalized retardates in a summer day camp. The subjects were all boys between the ages of 9 and 16. Camp counselors were told that certain boys were "bloomers." Observations of the amount of help given to the boys were made. There were no expectancy effects on IQ (TOGA); however, observations indicated that "bloomers" received less attention than "non-bloomers." The reasons for the differential attention were not investigated or known.

Pellegrini and Hicks (1972), administered the WISC similarities subtest and the Peabody Picture Vocabulary Test to children in a special tutorial project. The experimental sample consisted of 44 tutor-child pairs who were randomly assigned to high, average, or low expectancy condition. No significant effects were found on either measure.

Four studies have been reported which used some achievement index as the criterion measure of expectancy effects. Meichenbaum, Bowers, and Ross (1969), demonstrated the self-fulfilling prophecy effect using 14 female adolescent juvenile offenders. Six of these girls were designated "bloomers" and, because of time factors, IQ measures were not taken. Instead, subjective and objective exams prepared by teachers served as dependent variables. The experimental Ss scored significantly higher on the objective exam and demonstrated superior class-appropriate behavior. Teachers, at the beginning of the experi-

experiment, rated all girls as "good" or "poor" students. The six experimental girls were chosen on the basis of the teachers' rankings. Three "good" students and three "poor" students were supposedly going to "bloom" on the basis of a test which was, in fact, never given. While the teachers were originally skeptical of the three "poor" students, they were soon pointing to behaviors that indicated the girls actually did have academic potential, and they eventually concluded that the test to designate intellectual bloomers was a good test.

Schrank (1968, 1970) reported two successful expectancy studies. In each case, college freshmen were randomly assigned to a freshman mathematics course. Course instructors assumed, in the first study, that class assignment was based on ability level since such assignment had been routinely made in the past. The final course grade averages ranked the students from highest to lowest over five sections exactly as if the students had been assigned according to ability level. In the second study, instructors were told that the students were randomly assigned and, consequently, no significant differences were found among the sections using final grade averages as the criterion measure. Although some methodological difficulties are apparent, eg., use of repeated t-tests, these studies indicate the utility of further study.

A third study involving expectancy effects on achievement was reported by Palardy (1969). This investigation looked at the effects of existing teacher expectations on the probable success of learning to read. Specifically, Palardy gathered teachers' opinions on whether first grade girls or boys were more

likely to learn to read during the year. Two experimental groups were established using teachers' opinions as a categorizing scheme. One experimental group was comprised of five teachers who thought that girls would be more successful in learning to read. They were matched with the second group which was composed of five teachers who held that boys and girls had an equal probability of success. Using IQ as a covariate, reading achievement scores on the Stanford Achievement Test were significantly lower for boys in the unequal probability group than either group of girls or the boys in the equal probability group.

Beez (1970) reports a relatively superficial study supportative of expectancy effects on achievement. In this investigation, 60 graduate students served as tutors to 60 preschool children. The tutors were given bogus IQ rankings and the criteria measures were the number of symbols learned in a 15 minute session and the tutors' post hoc ratings of their students. Tutees listed as "high ability" learned more symbols and were rated higher than tutees listed as"low ability." It seems obvious, however, that since the tutoring situation is substantially different from the regular classroom system, and since the tutoring time consisted of a single 15 minute session, the results are, at best, indicative of expectancy effects.

In addition to the above noted studies concerning primarily cognitive effects of the self-fulfilling prophecy, a growing body of related literature discussing more affective effects is evident. These studies concern such topics as teacher-

pupil interactions, personality variables, and demographic variables. Jacobs (1970), for example, investigated expectancy effects using a sociometric measure as a dependent variable in 14 elementary school classrooms. Twenty percent of the students were identified as potential "stars." Posttesting after 10 weeks revealed no increases in peer acceptance, although a positive correlation was obtained between measured change and teachers' perceptions of sociometric change.

Haberman (1970) tried to induce expectancy effects between 120 student teachers and their cooperative teachers. Part of the student teachers were given high expectancies for their cooperative teachers and their cooperating teachers were given high expectancies for their student teachers. A control group was devised of those teachers given no information. After a semester of student teaching, no differences in rating were obtained between the experimental group and the control group. Haberman suggests that teachers may have clearly defined objectives and are therefore not susceptible to expectancy inducements.

Rothbart, Dalfen, and Barrett (1971) used micro-teaching as a vehicle to investigate teacher-pupil interactions. Thirteen female undergraduates and 52 high school students participated in a 30 minute literature discussion. Two of the four students in each micro-teaching group were labeled as "high academic potential" and two were labeled as having "low academic potential." Teachers tended to interact more with the "bright" students, and these students talked more ($p>.06$). In post hoc ratings, the teachers agreed that the "brighter"

students had greater potential. There were no significant differences in the number of positive reinforcements given to the two groups.

Another study using micro-teaching was conducted by Rubovits and Maehr (1970). Again, female undergraduates (N=26) served as teachers. The subjects were 104 sixth and seventh grade students who had been randomly labeled as "gifted" or "regular" on a seating chart given to each micro-teacher. Children labeled "gifted" received more praise and were asked more questions when compared to those labeled "regular."

Brown(1970) tried to induce expectancy effects using fake psychological reports. The reports included SES, intelligence, and personality factors. Ten teachers-in-training tutored 80 first grade children in a paired-associate task involving a list of states and their associated capitals. The false IQ scores were significantly related to the number of items the tutors attempted to teach; however, expectancy ratings of the teachers were in the opposite direction of the expectancy hypothesis. Furthermore, SES and personality variables had no significant relationship to the number of pairs learned.

Three studies are available in the literature which make use of existing teacher expectations and the subsequent pattern of teacher-pupil interactions. Willis (1970) had five special education teachers rank their classes of eight students each according to a criteria of "most efficient learners" to "least efficient learners." Observation revealed that the

teachers tended to ignore comments from less efficient learners and tended to give more verbal responses to comments of more efficient learners.

Good (1970) and Brophy and Good (1970) report findings relevant to teacher-pupil interactions. These investigators also had teachers rank their students according to achievement. Four regular first grade classrooms were used in each study. Three males and three females from the high achievement end of the teachers' ranking and three males and three females from the low end of the teachers' ranking were selected for observation. Good (1970) found that teachers interacted significantly more often with those ranked as high achievers. Good and Brophy found that high achievers initiate more interactions with the teacher than those ranked as low achievers. While the actual number of interactions with the teacher did not differ significantly between the two groups, high achievers seemed to have qualitatively different interaction. Criterion measures assessing achievement were found to favor the high achiever group. Good and Brophy note:

> The data show that teachers consistently favored the highs over the lows in demanding and reinforcing quality performance. Despite the fact that the highs gave more correct answers than did the lows, they were more frequently praised when correct, and less frequently criticized when incorrect or unable to respond. Furthermore, the teachers were more persistent in eliciting the responses from the highs than they were from the lows. . . (p. 372).

Two experiments, more indirectly related to classroom expectancy effects, have been reported. Cohen (1966) had 256 college education majors score subjective learning readiness

tests for hypothetical students. Bogus information about IQ and reading group placement was also provided. These teachers-in-training gave higher scores to "brighter" students. Similarly, Simon(1969) found that college students scored the vocabulary subtest of the WISC significantly higher for those subjects labeled "above average" in intelligence when compared with those labeled "below average" in intelligence. Actual WISC protocols were used for the vocabulary scoring although the labeling was randomly assigned.

A review of the literature revealed two more relevant studies. These studies, by Rist (1970) and Seaver (1971), are relatively unique and suggestive of alternative methods of studying the educational aspects of the self-fulfilling prophecy. Rist, a social anthropologist, conducted an interesting longitudinal study. Rist observed for one and one half hours twice a week in a kindergarten classroom of a ghetto school. He then observed the same children in first grade and for one semester in second grade. A carefully written observational record revealed some distressing, yet interesting, phenomena.

The kindergarten teacher, having access to information concerning the welfare status of each child's parents, placed the children at three different tables after eight days of classes. No test scores or other academic indicators were available. Differential treatment followed. The group at the first table, who most approximated a middle-class cleanliness model, received most of the teacher's time and attention. Indeed, the teacher began writing on the blackboard in such a manner that only those at the first table could see. The

children at the first table ran errands, were hall monitors, etc., and they even began to treat the children at the other two tables with the same distain as their teacher. By the end of the year, the only interaction children at the third table had had with the teacher was when she told them to sit down and/or keep quiet. Furthermore, the children at the third table began to treat each other in a hostile, unfriendly fashion, eg., calling each other stupid, dumb, etc. Because of this differential treatment, Rist argues, the children at the third table had not finished their pre-reading book by the end of the school year. Since these children had not been taught, they naturally scored low on the end of the year standardized achievement test. However, IQ tests given at the end of kindergarten revealed no significant differences among the three tables. Subsequently, the first grade teacher, who not only had subjective information, but also objective indicators in the form of test scores, assigned the table three children to the slow group with little positive expectation. This "caste system" was maintained throughout first and second grade.

The special significance of this study, aside from revealing some of the dreadful experiences some children are subjected to in schools, is that it indicates the overwhelming importance of non-academic factors in establishing teacher expectancies. The kindergarten teacher in this study started the cycle of the self-fulfilling prophecy in motion on the basis of whether a child's father was living at home, whether the family was on welfare, whether the child smelled because of uncleanliness, and a variety of other factors which are all irrelevant to the

child's academic potential. With this cultural factor impinging on the situation, attempts to change teacher expectancies on the basis of bogus IQ scores may well be an impossible task.

Seaver (1971) investigated the effects of naturally occuring expectancies by analyzing school permanent record data. In one sub-study, Seaver attempted a lagged correlation study in an attempt to predict grade point averages in grades 2-5 on the basis of IQ scores (Primary Mental Abilities) obtained in grade one. Results were uninterpretable probably because of the unreliability of grade one IQ test scores. More interesting were his findings regarding the expectancy effects caused by a teacher having previously had an older sibling of a child she currently had in her classroom. Records from two elementary schools revealed 79 pairs of siblings. A determination of whether the same teacher had taught one or both siblings was made. The older siblings were categorized as "good" or "bad" students by independent judges on the basis of first grade IQ scores, Stanford Achievement Test scores, and grade point averages. Younger siblings were then compared to older siblings on the basis of achievement test scores and grade point averages. Data analysis indicated that younger siblings of "good" students received higher grades if assigned to the same teacher when compared with those assigned to a different teacher. Also, younger siblings of "bad" students received significantly higher grades if assigned to a different teacher when compared with those assigned to the same teacher.

Again, like the Rist study, Seaver's work illustrates expectancy effects in naturally occuring situations without

the use of experimental manipulation.

C. Review and Synthesis:

The relatively exhaustive literature review has provided answers to some questions but raised other issues. Critiques of the Rosenthal research on experimenter bias and teacher expectancy have demonstrated that few, if any, conclusions can be drawn from his research. Methodological errors and conceptual errors of the kind suggested by Barber and Silver and Elashoff and Snow, among others, indicate that the research generated by Rosenthal and his co-workers is generally untrustworthy. The wide publication and use of the Rosenthal and Jacobson findings is apparently unjustified without further documentation.

Approximately ten other attempts to raise IQ scores through expectancy effects have been reported. In no case did the findings support the Oak School Experiment. Although methodological errors are also apparent in some of the replication studies, it seems appropriate to conclude that experimental manipulation of teacher expectations has no effect on IQ measures.

The research using achievement gains as dependent variables suggests that, under naturally occuring conditions or under strong experimental manipulation, teacher expectancies probably have some effect on student achievement under certain circumstances. Rist's findings suggest that teacher expectancy effects achievement in so far as the teacher refuses to interact with and teach certain children because of culturally established expectations.

Further, it seems fair to conclude that teacher-pupil interactions are affected by the expectations of those involved in the interactions. It seems clear that the number and quality of interactions between teachers and students are partially dependent on th. definition of the situation. In sum, the teacher expectation effect tends to occur in naturalistic situations and most prevalently in teacher-pupil interactions. Finally, since teachers' expectancies are based on non-cognitive factors, changes in cognitive information are not likely to alter the original expectations.

The only recent discussion, available in the literature, of the role of expectancy to larger educational issues is an article by Jeremy Finn (1972). Finn's distinction between two "kinds" of expectation is worth noting:

> It is the anticipation that shapes the manifestation of expectations. And it is anticipation that distinguishes expectations from hopes and desires, as well as from aspirations. While the concept aspirations implies some striving toward a desired goal, expectations incorporate an additional estimation of reality factors. That is, expectations imply the anticipation of the behavior most likely to actually occur, given the individual and circumstances (p. 390).

Finn uses the analogy of ghetto parents aspiring to have their children enter the professions but offering them little, if any, achievement oriented aid on which to fulfill expectations. He further argues that Rosenthal and Jacobson employ the concept of teacher expectancy in a similar fashion. That is, "They have attempted, however effectively, to set the teachers' estimates of the actual cognitive level at which the pupils would operate. The success of the experiment was a direct function of the

extent to which the contrived test results did in fact produce this effect (p. 391)."

It appears then, that Finn interprets the Rosenthal and Jacobson study as trying to influence the aspirations or desires that the teachers hold for their students without directly influencing the expectations (using Finn's dicotomy) that teachers hold for their students. Further, it could then be argued that the efforts to induce teacher expectancy via the Rosenthal model is a realistically ineffective approach if the intended effect is substantive behavior change.

CHAPTER 2

METHODS AND PROCEDURES FOR THE PRESENT EXPERIMENT

Although the current investigation was devised and underway before many of the critical analyses of the Rosenthal and Jacobson study were available in the literature, most of the methodological flaws in the Oak School Experiment, and subsequent replications, were avoided. Furthermore, the design permitted nearly all of the facets of teacher expectations to be tested.

Three general problems were considered in generating testable hypotheses: 1) Does experimental manipulation of teachers' expectations lead to increases in ability and achievement scores?; 2) If any score increases do occur, are they a function of the self-fulfilling prophecy or simply the result of different instructional programs?; and, 3) If any increases in scores do occur, and, if they are the result of a self-fulfilling prophecy, can we identify some of the processes involved?.

Ample evidence has been presented in Chapter 1 to suggest that the results reported by Rosenthal and Jacobson may well be spurious. It therefore seems clear that a methodologically sound replication is warranted. Since a major portion of this experiment and the entire notion of teacher expectancy hinges on increases in criterion measures as a result of experimental manipulation of expectancy, clear and unimpeachable increases in achievement and ability scores must be demonstrated.

The second problem relates to the possibility that the teacher may engage in different modes of instruction for those pupils who have been specifically classified. That is, if teachers expect a particular pupil to do especially well, he may place him in an instructional group which has better pupils, better instructional materials, and possibly different modes of instruction Pupils who had not been marked as deserving special consideration may not be given this attention.

In short, the gains in IQ and achievement which Rosenthal and Jacobson interpreted as expectancy effects caused by some subtle communication system may well have been the result of placing the experimental students into one instructional group rather than into another. This rationale suggests that, if there is a subtle communication system involved in the self-fulfilling prophecy, it may not be as important as the original instructional placement of the students by the teacher. The interpretation posited here may help explain why teachers could recall or recognize the names of only a few of the "bloomers" in the Oak School Experiment (Rosenthal and Jacobson, 1968).

Furthermore, Rosenthal and Jacobson report that experimental Ss in the middle range of ability made greater gains in reading achievement than either the low ability group or the high ability group. Assuming that original instructional placement is more important than the subtle communciation system, this finding suggests that teachers more readily believe that they could mistakenly place an average pupil in the wrong group but not a very competent pupil or a very incompetent pupil.

The third problem to be investigated is under the rubric of the process of the self-fulfilling prophecy and concerns the pupils' and the teachers' perceptions of expectancy. That is, are the expectancies conscious on the part of those involved? Specifically, (a) do the teachers have the expectations during and at the end of the treatment period; (b) do the matched pupils differ from their controls in their perceptions of what the teachers expect for them and for their matched counterpart; and (c) do the other pupils in the class recognize a different communication system between the teacher and the experimental subjects and the teacher and the matched control subjects? In other words, pupils must perceive that the teacher expects different achievement and behavior among them before a pupil can make a response contingent on the teachers' expectations. Therefore, teachers' expectations must be revealed in some form so that pupils can perceive differential treatment.

A review of previous research did not yield information directly concerned with pupils' abilities to perceive teacher expectations and differential treatment. However, indirect support can be taken from Flanagan and Havumaki (1960) who demonstrated that 10th grade pupils praised more often by the experimenter were chosen more frequently on a sociogram. Beck (1964) reported the development of a pupil perception of teacher behavior questionnaire, thereby demonstrating that intermediate level elementary school children could, and do, perceive at least eleven specific teacher characteristics. Lippitt and Gold (1959), and Fox, Lippitt and Schmuck (1964) found that

elementary school children perceived seventeen specific characteristics of their peers and that students utilized these perceptions in conceptualizing the power structure of the classroom. Furthermore, these perceptions remained consistent throughout the school year.

In sum, it appears that elementary school children make accurate perceptions of subtle affective and cognitive behaviors of their teachers and their peers. Furthermore, it seems reasonable that pupils can perceive differential expectations of the teacher and that these perceptions may be a vital link in whatever communication system may be involved in the self-fulfilling prophecy process.

Within the framework of these general problems, seven specific hypotheses were tested:

1. Experimental Ss will show significantly better performance on the ability posttest than the matched control Ss.
2. Experimental Ss will show significantly better performance on the achievement posttest than the matched control Ss.
3. When compared with their matched control subjects, experimental Ss in the middle ranges of ability will make significantly greater improvement in achievement and ability than either the high ability experimental Ss or the low ability experimental Ss.
4. Experimental Ss will receive significantly different instruction from that of the matched control Ss as measured by group classification and ratings by

independent observers.

5. Experimental Ss will report more positive, supportive interactions with their teachers on a questionnaire measuring pupils' perceptions of teacher behavior at the end of the school year as compared with the matched control Ss. Ratings of independent observers will support the responses of the experimental Ss.
6. Control Ss will report more positive, supportive interactions between their matched experimental counterparts and the teacher on a questionnaire measuring pupils' perceptions of teacher behavior at the end of the school year as compared with their own interactions with the teacher. Ratings of independent observers will support the responses of the control Ss.
7. Pupils who are not in the experimental group or the control group will report, on a questionnaire measuring pupils' perceptions of teacher behavior, significantly different pupil-teacher interactions between experimental Ss and the teacher as compared with the interactions of matched control Ss and the teacher.

Procedure:

1. Selection of Subjects and Experimental Design

All students in twenty-four classrooms, grades 1-6 (four classrooms at each grade level) were administered the Metropolitan Analysis of Learning Potential and the Metropolitan

on their achievement test scores and their potential scores. Classroom teachers were asked to rank their students according to how much achievement growth they expected from them during the year. Three experimental pairs were then selected from each classroom, one pair to represent each third of the distributions. That is, one experimental pair was selected if each member had an achievement score in the top third of the achievement distribution, if each had a potential score in the top third of the ability distribution, and if each had been ranked by his teacher in the top third of her achievement growth prediction distribution, etc. One further qualification for sampling was that the pairs were matched by sex. Furthermore, there was an equal number of males and females at each grade level, but not for each class in the sample. After the experimental pairs had been selected, one member of each pair was randomly assigned to the treatment group with the other member becoming part of the control group. Selection of Ss and implementation of the study was completed within the first four weeks of the fall term.

Schematically, the design is illustrated by Table I. Thus, there were 12 experimental Ss and 12 control Ss at each of the six grade levels, or a total of 72 experimental subjects matched with 72 control subjects. Classrooms were selected

---

[1]The first grade subjects found the Metropolitan Achievement Test too difficult and frustrating. Therefore, first grade Ss have scores only on the Analysis of Learning Potential.

TABLE I

EXPERIMENTAL DESIGN

GRADE	CLASSROOM	SEX	ABILITY GROUP	MATCHED PAIRS	EXPECTANCY TREATMENT	
					EXPER.	CONTROL
1	1	M	H	1	X*	X
		F	A	2	X	X
		M	L	3	X	X
	2	F	H	4	X	X
		M	A	5	X	X
		F	L	6	X	X
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
6	24	F	H	70	X	X
		M	A	71	X	X
		F	L	72	X	X

*X refers to the appropriate dependent variables under consideration in the study.

from two school districts in Central New York State on the basis of size and geographical location.

A questionnaire to discern pupils' perceptions of teachers' differential expectations and treatment of students was devised and tested for feasibility in classrooms not to be used in the study prior to the experiment. The instrument was administered to all students in the 24 participating classrooms at the beginning of the school year. Trained observers made regular visits to selected classrooms, and using the observation scale devised by Rist (1971), (see Appendix I), served as an added measure of pupil perception.

It has been hypothesized that instructional grouping of students may be more important than the communication system between the students and teacher. Therefore, a documentation of the grouping of experimental and control subjects was made.

2. Inducing Teacher Expectancy

Rosenthal and Jacobson (1968) did not seem overly concerned with the process by which teacher expectancy was produced. In fact, one could question whether any change in expectancy was induced by noting that the teachers involved in the Oak School Experiment were unable to recall or recognize most of the experimental subjects after the experiment had been concluded. It seems necessary to be concerned with the specific mechanisms by which expectancy is induced and to make the expectancy inducement as strong as possible. That is, to maximize any expectancy effects, maximum expectancy inducement ought to be used. The following procedure was used in the proposed study

because of the widespread popularization of the Rosenthal and Jacobson study.

It was noted earlier that each classroom teacher would be asked to rank his pupils according to how much he expected them to grow in achievement. After the experimental subjects were selected, the investigator had a conference with each teacher for the supposed purpose of discussing the differences between his ranking and test scores. Assuming that teachers, in general, distrust standardized test scores (see for example, Jackson, 1968, pp. 123-26), teachers were told that further research into this problem was being done in an effort to determine whether or not test makers should develop different kinds of standardized tests. Using this as a ploy, the investigator pointed out to the teacher, during the conference, that the scores obtained in the two tests given indicated that he had seriously underestimated the achievement potential of several of his pupils. With this in mind, we told him that we would like his cooperation in trying to find out whether the test estimates were accurate. The investigator then informed him that three pupils in his classroom, as well as three pupils in each of several other classrooms, had been singled out for close study. He was then given the names of the three experimental subjects and an expectancy inducement. That is, it was pointed out that while he may have ranked subject one as likely to grow, our tests indicated that this student should grow or improve much more rapidly than even he had first anticipated. He was then told that subjects 2 and 3, representing the middle third and the bottom third of his ranking,

ought to improve much more than he expected. In short, he was strongly induced to believe that these three students were going to make exceptional gains during the course of the year.

Finally, teachers were informed that observers would make regular visits to his classroom to see if the responses and school work which the experimental Ss produced in the classroom were consistent with responses that they made on the tests. It was also anticipated that the regular visits of these observers would serve as a reminder to the teacher and thereby reinforce the expectancy.

During the entire school year, one of three observers made weekly observations of approximately 45 minutes each in each of the twenty-four classrooms. Observers used the nine item teacher-pupil interaction scale developed by Rist (1971). Observers also noted changes in grouping, and other informal observations about the general atmosphere and operation of the classroom.

At the end of the school year, the 72 experimental Ss and the 72 control Ss were re-administered the Metropolitan Achievement Test and the Metropolitan Analysis of Learning Potential. The pupil perecption of teacher behavior questionnaire, developed by the investigator, was re-administered to all of the children in all twenty-four classrooms. After all posttesting was completed, the teachers were informed of the actual purpose of the experiment.

CHAPTER III

RESULTS AND DISCUSSION

The effects of the experimental manipulation of teacher expectancy was tested on eight dependent variables. The criterion measures covered a variety of academic and social factors. Of particular concern for each dependent variable is the comparison between experimental and control subjects and interactions involving the treatment factor. Differences were tested for significance by analysis of variance. Matched pairs within grade - ability level - treatment (d.f. =72) served as the error term for all comparisons. The two measures of intellectual or achievement growth were the gains in scores on the Metropolitan Analysis of Learning Potential and the Metropolitan Achievement Test. Since exact matching of test scores for the matched pairs was not always possible because of other matching criteria, the pre-test scores on these two instruments were treated as covariates in all analyses. It should also be noted that an exhaustive review of available computor programs did not yield a program which would analyze all the data and features of the present design in a single analysis. Thus, two separate analyses were done. The first encompassed grades 1-4 inclusive and the second encompassed grades 3-6 inclusive. Of the available options, it was concluded that retaining the two covariates was more valuable than analyzing all six grades without controlling for pretest differences. Therefore, achievement gains served as the

first dependent variable and ability gains served as the second criterion measure with the pretest scores on each measure being treated as a covariate.

The third variable tested is termed positive interaction. This data was derived from observations of pupil-teacher interactions. Three levels of positive interactions--verbal, non-verbal, and physical--were recorded and combined into a positive interaction score for each control subject and each experimental subject.

The fourth dependent variable, neutral interaction, was derived in the same manner as variable three except that interactions defined as neutral were recorded. Variable number five, negative interaction, matches variables three and four except that negative interactions, including verbal, non-verbal, and physical, were recorded. The sixth variable was the combined interaction score for each Ss. All interactions, from all three modes of interaction (physical, non-verbal, and verbal), were added into a single score.

The observation scheme used was developed and reported by Rist (1970). A copy of the observational record and a description of the behaviors covered by each category is presented in Appendix I. The clearness of the category parameters, together with pre-observation discussions, led to consistent recording of interactions.

The final two variables, sociometric-pre and sociometric-post, were used in testing Hypotheses 5, 6 and 7. The socio-

metric instrument used to generate this data was developed by the writer and is presented in Appendix II. Also presented in Appendix II is the results of a factor analysis (principal component method) of the twenty-three item scale. The scores used in the data analysis are the sum of the number of times each control subject and each experimental subject was chosen by anyone in his classroom on the thirteen items which clustered on the principal component.

The results of data analysis is presented in the following tables. The findings in Tables 2 and 3 show three, relatively trivial, significant F-ratios. These differences on Variables 1 and 2 in Table 2 and on Variable 2 in Table 3, indicate that posttest scores on the achievement and ability measures were significantly greater than pretest scores for both the experimental Ss and the control Ss. These findings suggest the obvious; namely, that all Ss made significant achievement gains over the course of a year. The only interesting factor about these two tables is why no significant achievement differences, on Variable 1, were found for grades 3-6 even though significant gains in ability scores were found.

Analysis of data by ability level within grade level yielded only one significant F-ratio. This finding indicated differences in the number of pupil-teacher interactions for experimental and control Ss combined for grades 1-4. Further analysis revealed that, for grade 1 Ss, high ability students had more interactions with the teacher than did average ability students and more interactions than did their low ability peers.

TABLE 2

Univariate F-tests by Grade. Grades 1-4 for Eight Variables (Test of G). Experimental and Control Ss Combined.

Variables	F (3, 10)	P Less Than
1. Achievement Gains	21.36	.001
2. Ability Gains	18.45	.001
3. Positive Interaction	0.97	.444
4. Neutral Interaction	2.09	.165
5. Negative Interaction	0.47	.707
6. Total Interaction	2.74	.099
7. Sociometric Pre	0.92	.465
8. Sociometric Post	3.42	.061

TABLE 3

Univariate F-tests by Grade. Grades 3-6 for Eight Variables (Test of G). Experimental and Control Ss Combined.

Variables	F (3, 10)	P Less Than
1. Achievement Gains	1.82	.207
2. Ability Gains	6.03	.013
3. Positive Interaction	1.01	.430
4. Neutral Interaction	0.51	.687
5. Negative Interaction	0.43	.733
6. Total Interaction	0.81	.519
7. Sociometric Pre	0.46	.717
8. Sociometric Post	1.12	.387

TABLE 4

Univariate F-tests by Ability Level Within Grade Level for Eight Variables (Test of GA). Grades 1-4, Experimental and Control Ss Combined.

Variables	F (6, 22)	P Less Than
1. Achievement Gains	0.91	.504
2. Ability Gains	1.50	.226
3. Positive Interaction	1.28	.307
4. Neutral Interaction	0.63	.702
5. Negative Interaction	0.65	.691
6. Total Interaction	3.22	.020
7. Sociometric Pre	1.26	.315
8. Sociometric Post	1.02	.440

TABLE 5

Univariate F-tests by Ability Level Within Grade Level for Eight Variables (Test of GA). Grades 3-6, Experimental and Control Ss Combined.

Variables	F (6, 22)	P Less Than
1. Achievement Gains	0.97	.468
2. Ability Gains	1.00	.450
3. Positive Interaction	1.62	.188
4. Neutral Interaction	0.78	.595
5. Negative Interaction	0.71	.647
6. Total Interaction	1.02	.439
7. Sociometric Pre	0.56	.754
8. Sociometric Post	0.24	.958

This finding is in agreement with other findings reported in the literature (eg., Good and Brophy, 1970). For Grade 2 Ss, findings indicated that average ability students and low ability students had significantly more interactions with their teachers than did high ability students. In Grade 3, low ability students interacted more with their teachers than either average ability students or high ability students. These findings are incongruent with others previously reported. One possible explanation involves differences in instructional methods used with different ability grouos. That is, very often teachers would assign high ability students independent projects and then leave them to their own initiative while she spent more time in more directed activities with average and low ability students.

An inspection of scores for Grade 4 Ss revealed no significant differences in the number of pupil-teacher interactions for the various ability levels.

When the data was analyzed by ability level without regard to grade level, two significant F-ratios were found. Both of these differences involved the observed interactions between pupils and teachers. For Grades 1-4 (Table 6), differences were found in the total number of pupil-teacher interactions. Inspection of raw scores indicates that low ability students had the most interactions with their teachers, followed by average ability students, with high ability students having the fewest interactions with their teachers.

For Grades 3-6 (Table 7), significant differences involved the positive interaction variable. Further analysis revealed

TABLE 6

Univariate F-tests by Ability Level Across Classrooms for Eight Variables (Test of A). Grades 1-4, Experimental and Control Ss Combined.

Variable	F (2, 22)	P Less Than
1. Achievement Gains	1.06	.364
2. Ability Gains	0.02	.985
3. Positive Interaction	0.19	.827
4. Neutral Interaction	0.19	.828
5. Negative Interaction	1.62	.221
6. Total Interaction	4.75	.019
7. Sociometric Pre	0.57	.571
8. Sociometric Post	0.70	.506

TABLE 7

Univariate F-tests by Ability Level Across Classrooms for Eight Variables (Test of A). Grades 3-6, Experimental and Control Ss Combined.

Variables	F (2, 22)	P Less Than
1. Achievement Gains	0.68	.519
2. Ability Gains	1.83	.183
3. Positive Interaction	4.48	.023
4. Neutral Interaction	2.18	.136
5. Negative Interaction	0.62	.546
6. Total Interaction	1.17	.329
7. Sociometric Pre	0.03	.975
8. Sociometric Post	0.92	.749

that low ability students had the largest number of positive interactions with their teachers; average ability students had the smallest number of positive interactions with their teacher; and high ability students had more positive interactions with their teacher than the average ability students but fewer than the low ability students.

The explanation for these findings is conjecture. The independent study hypothesis offered above might also apply to these results. Another possible alternative is that the absolute number of interactions is an artifact of the observation time. While matched pairs were observed for an equal amount of time, it is conceivable that observation time across grade level and within ability group, but without regard to the matched pair criterion, may not have been equal. Thus, low ability students may have been involved with the teacher for a larger share of the observation time than either of the two groups. This possibility is particularly appropriate to the lower grades where a large amount of instructional time is spent in small group instruction.

Tables 8 and 9 present particularly relevant data analysis. Testing for the treatment effect, experimental manipulation of teacher expectancy, between experimental Ss and control Ss revealed no significant F-ratios. The failure to find any differences involving the treatment factor on any of the various academic and social variables suggests rejection of Hypothesis 1, Hypothesis 2, Hypothesis 5, Hypothesis 6, and Hypothesis 7. These results would be consistent with the majority of studies

TABLE 8

Univariate F-tests. Treatment Effects, Experimental Ss Versus Control Ss, by Grade for Eight Variables, Grades 1-4 (Test of GT).

Variables	F (3, 10)	P Less Than
1. Achievement Gains	0.73	.558
2. Ability Gains	0.50	.689
3. Positive Interaction	0.53	.669
4. Neutral Interaction	0.07	.975
5. Negative Interaction	2.09	.165
6. Total Interaction	0.70	.537
7. Sociometric Pre	3.35	.064
8. Sociometric Post	1.46	.285

TABLE 9

Univariate F-tests. Treatment Effects, Experimental Ss Versus Control Ss, by Grade for Eight Variables, Grades 3-6 (Test of GT).

Variables	F (3, 10)	P Less Than
1. Achievement Gains	1.52	.269
2. Ability Gains	2.25	.145
3. Positive Interaction	0.61	.624
4. Neutral Interaction	1.37	.307
5. Negative Interaction	1.30	.327
6. Total Interaction	0.80	.520
7. Sociometric Pre	0.69	.975
8. Sociometric Post	0.31	.820

discussed in the previous Chapter.

When treatment effects were tested without regard to grade level, five significant F-ratios were obtained. These positive findings also involved the observed pupil-teacher interaction. Inspection of the raw data indicates that experimental subjects had more positive and neutral interactions with their teachers than the control subjects in Grades 1-4 (Table 10), when grade level and ability level were not considered. An examination of the raw scores relevant to Table 11 (Grades 3-6), suggests that experimental Ss had more total interactions as well as more neutral interaction and more positive interactions with their teachers when compared with the control Ss. While these findings seem to support the general thesis of teacher expectancy effects, an accurate interpretation of these results is unlikely. Since the design utilized in this experiment was specifically developed to test within grade level differences, differences across grade levels cannot be accounted for. It is possible that the difference across grade levels is a function of between grade variance rather than between experimental Ss and control Ss variance. In any event, these findings are, at best, weakly supportive of Hypotheses 5, 6, and 7.

Tables 12 and 13 present the crucial data analysis. When treatment effects were tested within grade level and by ability, no significant differences were obtained between the matched pairs. Unlike Tables 10 and 11, which demonstrated some main effects, the failure to find interaction effects rules out the

TABLE 10

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, Across Grade Levels for Eight Variables, Grades 1-4 (Test of T).

Variables	F (1, 10)	P Less Than
1. Achievement Gains	0.66	.436
2. Ability Gains	0.00	.985
3. Positive Interaction	5.30	.044
4. Neutral Interaction	13.28	.005
5. Negative Interaction	2.84	.123
6. Total Interaction	0.51	.490
7. Sociometric Pre	1.99	.189
8. Sociometric Post	0.87	.373

TABLE 11

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, Across Grade Levels for Eight Variables, Grades 3-6 (Test of T).

Variables	F (1, 10)	P Less Than
1. Achievement Gains	0.000	.985
2. Ability Gains	0.004	.948
3. Positive Interaction	5.66	.039
4. Neutral Interaction	6.57	.028
5. Negative Interaction	0.81	.389
6. Total Interaction	9.07	.013
7. Sociometric Pre	0.02	.880
8. Sociometric Post	0.34	.572

TABLE 12

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, by Ability Level Within Grade Level for Eight Variables, Grades 1-4 (Test of GAT).

Variables	F (6, 22)	P Less Than
1. Achievement Gains	1.42	.252
2. Ability Gains	1.33	.286
3. Positive Interaction	1.96	.116
4. Neutral Interaction	1.87	.132
5. Negative Interaction	0.61	.718
6. Total Interaction	0.39	.878
7. Sociometric Pre	1.28	.307
8. Sociometric Post	0.76	.606

TABLE 13

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, by Ability Level Within Grade Level for Eight Variables, Grades 3-6 (Test GAT).

Variables	F (6, 22)	P Less Than
1. Achievement Gains	1.42	.252
2. Ability Gains	1.33	.286
3. Positive Interaction	1.96	.116
4. Neutral Interaction	1.87	.132
5. Negative Interaction	0.61	.718
6. Total Interaction	0.39	.878
7. Sociometric Pre	1.28	.307
8. Sociometric Post	0.76	.606

possibility of accepting any specific hypotheses.[1] The lack of significant interaction effects for the treatment factor, combined with the failure to find any significant grade-level-treatment interaction effects (see Tables 8 and 9), forces a rejection of all hypotheses except number 4. Since Hypothesis 4 concerns differences in instructional method rather than differences on any of the eight dependent variables, it has yet to be considered. Before considering data relevant to this hypothesis, some closure on the analysis of variance is necessary.

Two final tables present the results of an analysis of treatment effects for different ability levels without regard for grade level. A single significant F-ratio is reported in Table 14. Again, the observed interaction variable is involved. Analysis of raw scores indicate that low ability Ss had the largest number of positive interactions with their teachers, followed by the average ability group, and then the high ability group for Grades 1-4. An inspection of Table 15 reveals no corresponding findings for grades 3-6. While this two factor interpretation effect is more suggestive than the main effects reported in Tables 10 and 11, it is not amenable to specific interpretations. At best, it suggests that students across four grade levels

---

[1] Winer notes "A significant over-all F test on a main effect, for example, indicates that one or more of a multitude of possible comparison is significant. The specific comparisons which are built into the design or suggested by the theoretical basis of the experiment can and should be made individually, regardless of the outcome of the corresponding over-all F test." B.J. Winer, Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962, pg. 208.

TABLE 14

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, by Ability Level for Eight Variables, Grades 1-4 (Test AT).

Variables	F (2, 22)	P Less Than
1. Achievement Gains	0.44	.649
2. Ability Gains	0.13	.881
3. Positive Interaction	9.41	.001
4. Neutral Interaction	1.71	.204
5. Negative Interaction	0.80	.460
6. Total Interaction	2.57	.099
7. Sociometric Pre	1.47	.251
8. Sociometric Post	1.62	.220

TABLE 15

Univariate F-tests for Treatment Effects, Experimental Ss Versus Control Ss, by Ability Level for Eight Variables, Grades 3-6 (Test AT).

Variables	F (2, 22)	P Less Than
1. Achievement Gains	1.90	.173
2. Ability Gains	1.67	.212
3. Positive Interaction	0.37	.695
4. Neutral Interaction	1.09	.355
5. Negative Interaction	0.83	.449
6. Total Interaction	2.59	.098
7. Sociometric Pre	0.46	.635
8. Sociometric Post	0.47	.632

have more positive interactions with their teachers if they are singled out for special treatment. This interpretation must be tempered by the fact that grade level differences were not accounted for; therefore, a conclusive statement cannot be made.

Since no significant achievement gains or ability gains were found, Hypothesis 4, regarding differences in instructional methods for experimental Ss when compared to control Ss, has no criterion measures on which to be evaluated. However, careful documentation of within classroom instructional methods and ability grouping revealed no significant differences between the teaching of experimental Ss and the teaching of control Ss. ..hile differences between ability groups was noted (eg., more independent study for "brighter" students), there was no indication that matched pairs within a given ability level were treated differently. High ability control subjects participated in as many independent projects as did their experimental counterparts.

It should be stressed that these findings do not rule out the possibility that achievement differences reported by other investigators may have been the result of differences in instructional methods rather than exoerimental manipulation of teacher expectancies. The failure to find achievement differences in the present study makes it impossible to rule out the instructional methods hypothesis.

In sum, data analysis revealed that of 92 F tests computed, only 12 reached acceptable levels of significance. Of the 12

significant F-ratios, three, regarding over-all achievement gains, were relatively uninteresting. Of the remaining 9, 8 were generally uninterpretable main effects involving observed pupil-teacher interactions. One ability level-treatment interaction effect was reported in Table 14. This effect was not amenable to concise interpretation. Thus, all experimental hypotheses were rejected or, in the case of Hypothesis 4, rendered untestable.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The failure to find any significant effects of experimental manipulation of teacher expectations leads this investigator to concur with others in regarding the Rosenthal and Jacobson findings spurious. In a relatively well executed randomized block design, the failure to find interaction effects supporting the specific hypotheses for any of eight variables leads to the conclusion that teachers' expectations, and subsequent behaviors which may affect student achievement, cannot be altered by the simple process of showing teachers cognitively inconsistent test scores. That is, while teachers' expectations may well affect the social and achievement behaviors of students, the etiological factors are varied and complex. One inconsistent test score cannot dispel other social and academic factors which impinge upon the teacher and serve as the basis for expectations.

Although some investigators have reported gains in achievement test scores as a result of expectancy inducement, the majority of negative findings suggest that significant positive findings are either situation-specific or chance findings.

Since no significant differences in instructional methods for groups within classrooms was observed except presenting the same material at a slower pace, hypotheses concerning modes

of instruction cannot be evaluated. Furthermore, since grouping and regrouping was at a minimum in all classrooms, no basis for the assessment of the group placement hypothesis is available. Since no significant results in criterion measures were found, the effects of group placement and instructional methods could not be determined. It could be argued, however, that the lack of regrouping was one of the chief reasons that there were no increases in the achievement posttest. That is, if group placement is a major cause in making the teacher's expectancy come true, the failure of the teacher to regroup experimental subjects could have contributed to the lack of significant differences in achievement measures.

Based on the findings reported in this study, the other findings reported in post Rosenthal and Jacobson studies, and an analysis of theoretical positions regarding the self-fulfilling prophecy, it is the opinion of this writer that alternative modes of investigation be pursued. It seem clear that a teacher's expectations are founded on a variety of complex and inter-related factors; and it is naive to assume that showing a teacher one or two scores on a standardized achievement test will have any effect on his original expectations. It is recommended that observational studies, following Rist (1970) and historical studies, a la Seaver (1971) be undertaken in an effort to identify the various social, psychological, and academic factors utilized by teachers when they set expectations. While this writer still holds the original

claim that teacher expectations affect pupil behavior and achievement, attempts to experimentally manipulate or alter those expectations are premature until the basis for and the operations of teachers' expectancies is ascertained. Candidly, it seems unlikely that we can experimentally manipulate teachers' expectations if we do not know what factors or antecedents to manipulate.

Another avenue of investigation which should be pursued was suggested by Finn (1972). Quite correctly, Finn notes that the teacher is only one (perhaps minor) source of expectancy for the student. Parental expectations, peer expectations, and self-expectations are also impinging on the pupil's behavior and achievement. A wide variety of empirical questions are readily obvious. Is one source more potent than the others? What is the effect of incongruent expectancies from different sources? Does a change in the expectations of one source influence the child's perceptions of expectancy from other sources? Indeed, it is possible that the predominance of other sources of expectancy makes the teacher's expectations rather inconsequential.

In conclusion, it seems clear that the expectations of others play an influential role in the behavior of the individual, both in school and in the larger social arena. Furthermore, the basis for the expectations and the individual's perception of the various expectations are complex and generally virgin areas for investigation. Finally, it seems fair to conclude that investigations following the Rosenthal and Jacobson

model have demonstrated the model to be naive, premature, and founded on inadequate evidence. It is hoped that investigations following the ideas outlined above will be undertaken, and that these investigations will provide an understanding of a potential source of significant educational change and improvement.

APPENDIX I

Observation Schedule

1. Verbal supportive - "That's a very good job." "You are such a lovely girl." "Ky, but your work is so neat.

2. Verbal neutral - "Laura and Tom, let's open our books to page 34." "Ray, your pencil is on the floor." "Hal, do you have milk money today?"

3. Verbal control - "Lou, sit on that chair and shut up." "Curt, get up off that floor." "Ray and Laura, quit your talking."

4. Non-verbal supportive - Teacher nods her head at Rose. Teacher smiles at Liza. Teacher claps when Laura completes her problem on the board.

5. Non-verbal neutral - Teacher indicates with her arms that she wants Lilly and Shirley to move farther apart in the circle. Teacher motions to Joe and Tom that they should try to snap their fingers to stay in beat with the music.

6. Non-verbal control - Teacher frowns at Lena. Teacher shakes finger at Amy to quit tapping her pencil. Teacher motions with hand for Rose not to come to her desk.

7. Physical contact supportive - Teacher hugs Laura. Teacher places her arm around Mary as she talks to her. Teacher holds Trish's hand as she takes out a splinter.

8. Physical contact neutral - Teacher touches head of Nick as she walks past. Teacher leads Rena to new place on the circle.

9. Physical contact control - Teacher strikes Lou with stick. Teacher pushes Curt down to his chair. Teacher pushes Hal and Doug to the floor.

From Rist, R. 1970 (pp. 438-439).

APPENDIX II

Socio-metric Instrument.

Factor analysis of the instrument.

Final items used in the data analysis.

Name _____

WHO ARE THESE PEOPLE

DIRECTIONS:
Read each description below and write the name of the student(s) it describes in the space to the right. It is possible that some names will not appear in any of these descriptions. Also, it may be possible that your name will fit one or more of the descriptions.

1. Who is quiet but seems to be trying? _____
2. Whom do you think has good ideas? _____
3. Whom does the teacher think has good ideas? _____
4. Who likes to be told what to do? _____
5. Whom does the teacher smile at most? _____
6. Whom would you pick to see that the job gets done? _____
7. Whom does the teacher help most? _____
8. Whom does the teacher expect to do best? _____
9. Who gets a second chance if his (her) answer is wrong? _____
10. Who gets to be the group leader? _____
11. Who gets extra help from the teacher? _____
12. Who gets to be the teacher's helper first? _____
13. Who gets to do things first? _____
14. Who gets to sit next to the teacher? _____
15. Who talks to the teacher most? _____
16. Who is the best worker? _____
17. Who is the best pupil? _____
18. Who does the best work? _____
19. Who never gets into trouble? _____
20. Who is the smartest pupil? _____
21. Who reads the best? _____
22. Whom do you like best? _____
23. Who is your best friend? _____

ROTATED FACTOR MATRIX FOR SOCIO-METRIC SCALE

	FACTOR			
VARIABLE	1	2	3	4
1	.09812	.06720	.83251	.03434
2	.71384	.10497	-.00177	.19685
3	.79564	.03190	.02980	.19041
4	.10224	.72098	.17395	.22196
5	.63560	.12423	.31164	.11884
6	.56745	-.08707	.30898	.02926
7	.05553	.84236	.00549	-.02421
8	.65230	.15694	.23570	.09914
9	.10418	.78626	.00923	.00907
10	.60262	.04564	.16382	-.10464
11	.00264	.85711	.00555	-.06691
12	.68473	.05361	-.02827	-.00427
13	.69141	.02436	.16960	.12254
14	.44381	.35196	-.18846	.25485
15	.62119	.05183	.47770	.17575
16	.51814	-.00807	.61360	.22807
17	.69515	.08760	.35706	.20859
18	.38330	.04806	.76800	.14281
19	.67629	.02791	.36372	.18025
20	.72487	.07833	.23416	.12941
21	.22339	.06640	.10291	.88263
22	.28101	.01731	.19448	.82738

Note: Item 14 was eliminated before the factor analysis because it was situation soecific for many of the younger children. That is, they did not generalize beyond the immediate testing situation and consequently most children selected the child who was currently sitting next to the teacher.

ITEMS LOADING ON PRINCIPAL COMPONENT

FOR THE

WHO ARE THESE PEOPLE QUESTIONNAIRE

1. Whom do you think has good ideas?
2. Whom does the teacher think has good ideas?
3. Whom does the teacher smile at most?
4. Whom would you pick to see that the job gets done?
5. Whom does the teacher expect to do best?
6. Who gets to be the group leader?
7. Who gets to be the teacher's helper first?
8. Who gets to do things first?
9. Who talks to the teacher most?
10. Who is the best worker?
11. Who is the best pupil?
12. Who never gets into trouble?
13. Who is the smartest pupil?

REFERENCES

Anderson, D.F. and Rosenthal, R. "Some effects of interpersonal expectancy and social interaction on institutionalized retarded children." Proceeding of the 76th Annual Convention of the American Psychological Association, 3 (1968): 479-480.

Barber, T.X. "Invalid arguments, post mortem analyses, and the experimenter bias effect." Journal of Consulting and Clinical Psychology, 33 (1969): 11-14.

Barber, T.X. et al. "Five attempts to replicate the experimenter bias effect." Journal of Consulting and Clinical Psychology, 33 (1969): 1-6.

Barber, T.X. and Silver, M.J. "Fact, fiction, and the experimenter bias effect." Psychological Bulletin Monographs, 70 (6, Part 2) (1968): 1-29.

Barber, T.X. and Silver, M.J. "Pitfalls in data analysis and interpretation: A reply to Rosenthal." Psychological Bulletin Monograph, 70 (1968): 48-62.

Beck, ..R. "Pupil's perceptions of teacher merit: A factor analysis of five postulated dimensions." Dissertation Abstracts, 1964.

Beez, .V. "Influence of biased psychological reports on teacher behavior and pupil performance." Learning in Social ettings, eds., M.B. Miles and W.W. Charters, Jr., Boston: Allyn and Bacon, Inc., 1970.

Brophy, J. and Good, T.L. "Teachers' communication of differential expectations for children's classroom performance: Some behavioral data." Journal of Educational Psychology, 61 (1970): 356-374.

Brown, W.E. "The influence of student information on the formulation of teacher expectancy." Dissertation Abstracts, 30 (1970): 4822-A.

Burnham, R. "Experimenter bias and lesion labeling." Unpublished paper, Purdue University, 1966.

Claiborn, W.L. "Expectancy effects in the classroom: A failure to replicate." Journal of Educational Psychology, 60 (1969): 377-383.

Cohen, L.S. "An experimental manipulation of the halo effect." Doctoral dissertation, Stanford University, 1966.

Conn, L., Edwards, C., Rosenthal, R., and Crowne, D. "Emotion perception and response to teacher expectancy in elementary school children." Psychological Reports, 22 (1968): 27-34.

Cordaro, L. and Ison, J.R. "Observer bias in classical conditioning of the planarian." Psychological Reports, 13 (1963): 787-789.

Elashoff, J. and Snow, R. Pygmalion Reconsidered. Worthington, Ohio: Chas. Jones Publishing Co., 1971.

Evans, J. and Rosenthal, R. "Interpersonal self-fulfilling prophecies: Further extrapolation from the laboratory to the classroom." Proceedings of the 77th Annual Convention of the American Psychological Association, 4 (Part 1) (1969): 371-372.

Finn, J. "Expectations and the educational envoronment." Review of Educational Research, 42 (1972): 387-410.

Flanagan, N. and Havumaki, S. "The effects of teacher-oupil contacts involving praise on the sociometric choices of students." Journal of Educational Psychology, 51 (1960): 65-68.

Fleming, E. and Anttonen, R. "Teacher expectancy, or my fair lady." American Educational Research Journal, 8 (1971): 241-252.

Flowers, C. "Effects of an arbitrary accelerated group placement on the tested academic achievement of educationally disadvantaged students." Dissertation Abstracts, 27 (1966): 991-A.

Fox, R., Lippitt, R., and Schmuck, R. Pupil-Teacher Adjustment and Mutual Adaptation in Creating Classroom Learning Environments. Final Report, Cooperative Research Project, No. 1167, University of Michigan, 1964.

Goldsmith, J.S. and Fry, E. "The test of a high expectancy prediction on reading achievement and IQ of students in grade 10 (or, Pygmalion in puberty)." Submitted to the American Educational Research Journal, 1970.

Good, T.L. "Which pupils do teachers call on?" The Elementary School Journal, 70 (1970): 190-198.

Haberman, M. "The relationship of bogus expectations to success in student teaching." The Journal of Teacher Education, 21 (1970): 69-72.

Ingraham, L. and Harrington, G. "Psychology of the scientist: XVI. Experience of E as a variable in reducing experimente. bias." Psychological Reports, 19 (1966): 455-461.

Jackson, P. Life in Classrooms. New York: Holt, Rinehart and Winston, 1968.

Jacobs, J.F. "Teacher expectancies: Their effect on peer acceptance." Paper presented at Annual Meeting, American Educational Research Association, Minneapolis, 1970.

José, J. and Cody, M. "Teacher-pupil interactions as it relates to attempted changes in teacher expectancy of academic ability and achievement." American Educational Research Journal, 8 (1971): 39-49.

Kester, S. "The communication of teacher expectations and their effects on the achievement and attitude of secondary school students." American Educational Research Journal, 66 (1972): 51-55.

Lippitt, L. and Gold, M. "Classroom social structure as a mental health problem." Journal of Social Issues, 15 (1959): 40-58.

Meichenbaum, D., Bowers, K., and Ross. "A behavioral analysis of teacher expectancy effect." Journal of Personality and Social Psychology, 13 (1969): 306-316.

Palardy, J. "What teachers believe, what children achieve." Elementary School Journal, 69 (1969): 370-374.

Pellegrini, R. and Hicks, R. "Prophecy effects and tutorial instructions for the disadvantaged child." American Educational Research Journal, 9 (1972): 413-420.

Pitt, C. "An experimental study of the effects of teachers' knowledge or incorrect knowledge of pupil I.Q.'s on teachers' attitudes and practices and pupils' attitudes and achievement." Dissertation Abstracts, 16 (1956): 2387-2388.

Rist, R. "Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education." Harvard Educational Review, 40 (1970) 411-451.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. "Self-fulfilling prophecies in behavioral research and everyday life." Claremont Reading Conference Thirty-Second Yearbook, 1968, ed., M.A. Douglass, 16-33.

Rosenthal, R. and Fode, K. "The effect of experimenter bias on the performance of the albino rat." Behavioral Science, 8 (1963): 183-189.

Rosenthal, R. and Jacobson, L. "Teachers' expectancies: Determinants of pupils' I.Q. gains." Psychological Reports, 19 (1966): 115-118.

Rosenthal, R. and Jacobson, J. Pygmalion in the Classroom. New York: Holt, Rinehart and inston, Inc., 1968.

Rubovits, P. and Maehr, M. "Pygmalion analyzed toward an exploration of the Rosenthal-Jacobson findings." Paper presented at Annual Meeting, American Educational Research Association, Minneapolis, 1970.

Schrank, W.R. "Further study of the labeling effects of ability grouping." The Journal of Educational Research, 63 (1970): 358-360.

Seaver, W.B. "Effects of naturally induced teacher expectancies on the academic performance of pupils in primary grades." Doctoral Dissertation, University of Illinois, 1971.

Simon, K.E. "Expectancy effects in the scoring of vocabulary items: A study of scorer bias." Journal of Educational Measurement. 6 (1969): 159-164.

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

Willis, B.J. "The influence of teacher expectation on teachers' classroom interaction with selected children." Dissertation Abstracts, 30 (1970): 5072-A.