INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again – beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

A Xerox Education Company

73-2848

KERN, Joseph C., 1936-
EXPERIMENTER EFFECTS IN PUPIL TESTING.

The City University of New York, Ph.D., 1972
Psychology, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan

EXPERIMENTER EFFECTS IN PUPIL TESTING

By

Joseph C. Kern

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment for the degree of Doctor of Philosopy, The City University of New York.

1972

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May 31, 1972
date

Walter Weiss
Chairman of Examining Committee

May 31, 1972
date

Walter Weiss
Executive Officer

Prof. Morey Wantman

Prof. Walter Weiss

Prof. Harry Kaufman
Supervisory Committee

The City University of New York

Acknowledgment

I wish to acknowledge the help of Professor Walter Weiss to whom I am deeply grateful for his continued support. Gratitude is also expressed to Professor M. Wantman for his many hours of advice.

A special note of thanks is extended to the faculty and children of the third and fourth grades in the Glen Head Elementary school for their enthusiastic participation.

To my wife, Annette, I express my gratitude for her understanding and spirit of help.

Table of Contents

TABLES

I. Introduction: self-fulfilling prophecies

> If you treat an individual as he is, he will stay as he is, but if you treat him as if he were what he ought to be, he will become what he ought to be and could be.

This statement by Goethe has become the guiding principle behind the concept currently referred to as "the self-fulfilling prophecy". Robert Merton's (1948) formulation provided the stimulus for much of the research which followed in subsequent years. He employed the concept to explain the failure of banks as well as such large scale social and economic phenomena as racial and religious prejudice.

Since then, the concept has been adopted in other fields of inquiry. Within the area of psychotherapy, Frieda Fromm-Reichmann spoke of "iatrogenic psychiatric incurabilities" meaning that the therapist's own belief about the patient's prognosis might determine that prognosis. A formal investigation conducted by Goldstein (1962) found that the therapist's expectation was related to duration of psychotherapy.

The well known placebo effect is another example of the operation of self-fulfilling prophecies. In a review of the placebo literature, Shapiro (1960) notes that new drugs always seem most effective early in their career and lose some of their therapeutic potency with the passage of time. Fisher, Cole, Rickels and Uhlenjuth (1964) provide some experimental support for this assertion. In studying the effects of meprobroate, they trained one group of physicians to communicate grave doubts about the efficacy of the drug. The more

confident, enthusiastic physicians were more successful in keeping their patients in treatment.

Expectancy effects have also been found in survey research. For example, Wyatt and Campbell (1950) trained over 200 student interviewers for a public opinion survey dealing with the 1948 presidential campaign. Before interviewing began, the interviewers predicted the percentage distribution of responses on each of five questions. For four of the five questions, interviewers tended to obtain responses that were consistent with their predictions.

In short, from the fields of psychotherapy and medical and survey research, there is evidence that self-fulfilling prophecies play some role in determining behavioral outcomes.

II. Evidence from laboratory research through 1969*

Main Effect.- A comprehensive plan of research dealing with the effects of experimenter expectancy was begun by Rosenthal (1966) in the early 1950's. In his 1969 summary of the studies, he notes that the basic paradigm has been to establish two groups of experimenters and to generate for each a different expectation of their subject's responses. There have been 94 studies conducted on expectancy effects in seven

---

* A distinction is being made here between studies published prior to 1969 and the more recent ones published since then. Prior to 1969 many expectancy studies were simplistic in nature and were designed to demonstrate the pervasiveness of the phenomenon. Recent studies have been more elegant in design focusing on interactions, isolating the perameters of the effect and broadening the theoretical framework. Since they constitute a more sophisticated approach, they will be dealt with as a separate unit.

research areas: animal learning, human learning and ability, psychophysical judgments, reaction time, inkblot tests, laboratory interviews, and person perception. Of these 94 studies, 50% yielded a p value at the .10 level; 35% at the .05 level; 17% at the .01 level; and 22% at $<.001$ level.

On the basis of these studies, Rosenthal (1969) asserted that within a laboratory setting "the results of an experiment may be determined at least in part by the expectations of the experimenter". Morgan and King (1966), in their introductory psychology test, state that "the phenomenon (expectancy effects) is, apparently, reliable as well as general".

This conclusion has recently been challenged by Barber and Silver (1968). They critically analyzed 32 studies that attempted to demonstrate expectancy effects, and concluded that the reliability of the phenomenon has as yet not been clearly demonstrated.

According to their analysis, there were major deficiencies in 19 studies as follows: (1) An overall statistical analysis was not performed to exclude chance findings. (2) A large number of post hoc statistical tests were performed after the overall analysis had failed to reject the null hypothesis. (3) "Probability pyramiding" occurred in which there was a failure to take account of changing levels of significance when many statistical tests were performed on a single set of data. (4) P values below the conventional .05 level were accepted as substantiating the experimental hypothesis. (5) Data which were significantly opposite to the experimental hypothesis were not used in the statistical

analysis which ostensibly showed the experimenter bias effect. (6) A multitude of comparisons were made on overlapping data using individual t tests and Spearman RHO. The reviewers felt it would have been more appropriate to use a multivariate approach since there were many independent and many dependent variables.

Barber and Silver (1968) concluded that "the experimenter bias effect appears to be more difficult to demonstrate and less pervasive" than previously thought.

In a rebuttal to the above critique, Rosenthal (1968) notes that the 32 studies should be viewed as successive replications and runs of experiments rather than as 32 discrete formulations. As a consequence, he used Stouffer's method for combining the probabilities from a run of experiments and obtained "a combined p of less than one in a million-raised to-some-power-greater-than-two..." He therefore concluded that this analysis "provides us with little reassurance that the results of behavioral research are unaffected by our hypothesis or expectations".

Rosenthal then takes up the six dificiencies mentioned by Barber and Silver and offers his own rebuttal of each: (1) He has never performed any overall tests that perfectly "exclude chance findings" and is "not optimistic that such tests will be developed". (2) Statistical tests as such do not reject hypotheses--"only people do that and they differ in how they do it." He prefers to keep track of whatever p

values are obtained rather than adhere rigidly to a conventional accept-reject model. Furthermore, the post hoc comparisons actually derived from planned comparisons. (3) Most of the tests that could be criticized as probability pyramiding were made on a variety of subsidiary hypotheses, almost all of which had been examined several times in the past. He notes that the best cure for probability pyramiding is replication and that this is exactly what has been done. (4) The acceptance of p values at the .10 level of confidence is defended by Rosenthal who prefers continual replications and the revision of his judgment as new data come in. (5) Rosenthal simply rejects the charge that negative data were excluded from the statistical analysis. (6) He feels that a factor analytical or multivariate approach would be inappropriate, given the small N in most of the studies and that "in most studies most of the relationships tested by T's and RHO's were replications."

Rosenthal then notes that, for the main effect of expectancy, twenty of the thirty-five studies reviewed yielded "a two-tailed p of .10 and five of the fifteen 'negative' results showed the primary interaction to be significant at the .10 level (two-tailed)." Finally, he points out that Barber and Silver's review included less than half of the studies he conducted. His conclusion is that the effects of experimenter expectancy are robust, general, and serious enough to warrant the employment of procedures designed to permit their control.

In their rejoinder, Barber and Silver (1968) restate the six critical points noted above, feeling that they were not adequately answered by Rosenthal. They offer additional criticisms noting that post mortem tests were performed on non significant data and that invalid statistical procedures were used. After a review of Rosenthal's research to exemplify these criticisms, they end by restating their previous conclusion that the experimenter expectancy effect has not been clearly demonstrated.

Interpersonal Mediating Variables.-- In his 1966 volume and more extensively in his 1969 manuscript, Rosenthal has reviewed the pertinent literature on the potential determinants of expectancy effects.* First, he deals with the possibility that expectancy effects are basically due to methodological artifacts. Barber and Silver (1968) had suggested that cheating and/or recording errors could account for the main effect of expectancy in those studies where it was demonstrated. While agreeing that such errors can and do play a part within a single experiment, Rosenthal believes that this explanation alone is not sufficient to account for the effect.

In several studies where cheating and/or recording errors were clearly controlled, a significant main effect was still found. For example, Adair and Epstein (1967) who obtained an expectancy effect, had presented tape recorded instructions to the S's, who then recorded their own responses. Also, the responses of Johnson's (1967) subjects were recorded by an

* Table 1 (Appendix A) summarizes the studies to be described below.

electrical system that did the bookkeeping. Despite the tightness of his controls over cheating and recording errors, he obtained a very large main effect of experimenter expectancy. Rosenthal (1966) cites several other studies in which a similar procedure was used to isolate the effects of cheating and recording errors and reports that in many cases there was no evidence of the operation of these factors. In those studies where cheating or recording errors was found, the rates were too low to account for the results of the study.

Second, it has been suggested that operant conditioning may mediate the effect of experimenter expectancy. That is, the experimenter may unwittingly reinforce expected responses with a smile, a glance or some other signal. Such a model would suggest that the very first response of a given subject is not affected by the experimenter's expectancy and that later responses are more affected than earlier ones.

The accumulated evidence suggests that operant conditioning alone is insufficient to account for the occurrence of expectancy effects. Rosenthal (1966) described three studies that called for the experimenter to administer a photo rating task. One group was led to expect +5 and another group -5 ratings. Comparisons were made between ratings of the very first photo and all ten photos combined. For all three studies combined, the grand mean difference based on the first photo alone was 1.34 while that for all ten photos combined was 1.23.

Similarly, Masling (1965), in a study of examiner effects in influencing subjects' Rorschach responses, was unable to defend a reinforcement model. Weick (1965) used a set of twenty photos and induced expectations of +5 and -5 ratings for different groups of experimenters. Contrary to a reinforcement model prediction, the earliest responses showed the greatest effects of the induced expectations. Adair and Epstein (1967) also ruled out the operation of operant conditioning as a mediator in their research.*

Third, as a further attempt to identify the mechanism by which expectancy is communicated, an experimental strategy has been developed in which the visual and auditory channels are analyzed separately. Several studies point to the importance of the auditory channel in the communication of expectancy effects. In the experiment by Adair and Epstein (1967), sub-

* Although the reinforcement model does not serve as a general explanation of the "how" of the effect, it does shed some light on the "when" question. Rosenthal (1966) has summarized a body of data leading to the general conclusion that the experimenter expectancy effect occurs very early in the E-S interaction. The studies mentioned above suggest that, during the brief period when the E greets, seats, and instructs his subject, the expectation of the experimenter may already have been determined. A series of filmed experiments conducted by Rosenthal (1966) were rated by independent observers as to the expectation of the experimenter, as well as when they thought the expectation was transmitted. In many cases, the raters were able to identify accurately the expectation of the experimenter from his early moments of contact with the subject. Stevenson and Odom (1963) found that the sex of the experimenter affected the subject's performance even though the experimenter left the subject after giving instruction and before the subject began responding. Rosenthal (1966) suggests that "in future studies we must focus our attention on the brief pre-data collection phase of the experimental interaction in order to discover the 'how' of the communication of expectancies."

jects heard only the instructions that were tape recorded by experimenters who had been given different expectancies. The z associated with vocal expectancy alone was +1.64. Troffer and Tart (1964) had some trained hypnotists read standard passages to subjects. One group expected their subjects to be highly suggestible while another group expected theirs to be much less suggestible. The voices of the group expecting lower suggestibility scores were found to be significantly less convincing in their reading of the instructions to the subjects (z=+2.81).

Rosenthal and Fode (1936b) conducted an experiment employing a person perception task. This experiment used three groups of experimenters. Two groups were permitted access to both the visual and auditory channels but each of these groups had opposite expectations concerning the performance of their subjects. The third group of experimenters was allowed access only to the auditory channel; in addition, opposite expectations were induced for this group. Results showed that 47% of the total expectancy effect was obtained when subjects had access only to the auditory channel. Duncan and Rosenthal (1968) also supported the importance of the auditory channel. Finally, in those studies in which subjects had no access to auditory cues, little or no expectancy effect was found (Carlson and Hergenhahn, 1968; Moffatt, 1966).

Data are also available to demonstrate the importance of the visual channel. Zoble's work (1968), described above, found that visual cues were more effective in the mediation of expectancy

effects than were auditory cues. Those subjects who had access only to visual cues were affected by their experimenter's expectancy 75% as much as those subjects who had access to both information channels.

The above findings are difficult to interpret. Perhaps when subjects are deprived of either visual or auditory information, they focus heavily on the channel that is available to them. Such greater attention to the available channel might enable subjects to extract more information from it than they would ordinarily when both channels are operative.

Fourth, expectancy effects may be learned phenomena and learned in interaction with a series of research subjects. The experimenter could emit unprogrammed and unintended cues to his early subjects and as they respond in the expected direction this reinforcement could shape the pattern of the experimenter's unintended cues. To support such a proposition, it would have to be shown that experimenters are more successful in their unintentional influencing of subjects later, rather than earlier, in the sequence of subjects tested.

Rosenthal (1969) has summarized twelve studies bearing on this formulation and found that six support the hypothesis, one runs counter to the hypothesis, and five neither support nor refute the hypothesis. The overall directional z in support of a learning hypothesis is +3.06. In short, it would appear that it is among later seen subjects that expectancy effects are likely to be larger.

Canavan (1968) recently has attempted to isolate mediating

interpersonal factors in the communication of expectancy effects. She begins with the reward-cost orientation model provided by Thibaut and Kelley (1961). A reward-oriented person is one who pays attention to potential rewards in the environment while a cost-oriented person maintains his satisfaction by keeping his standards relatively low and by using his energy to avoid costs and failures. She suggests that the development of reward or cost orientation may be a kind of self-fulfilling prophecy which arises from the social labeling of successes and failures by "powerful others" in the environment. Hence, the reward-oriented individual would be expected to develop higher standards for performance than would the cost-oriented person, who would confine his aspirations to relatively low levels of outcomes in order to maintain a high probability of avoiding failure. This proposition was tested by initiating a "reward-orientation" cycle in one group and a "cost orientation" cycle in another. Subjects were 64 fifth and sixth graders in a suburban New York school who were assigned to one of four treatment groups.

Canavan reports that the group iven only the reward-success treatment responded in the predicted direction. Subjects experiencing only punishment-failure fell at the other extreme on all the measures, while subjects in the other two treatment groups fell between the extremes.

Organismic.--Summarizing the results of four studies investigating the effects of the sex of the experimenter in a person perception task, Rosenthal (1969) has concluded that when differences are found, males tend to show the greater

effects. However, the sex of the subject as well as the task involved are complicating factors. Johnson (1967), in a marble dropping experiment, found that there were greater expectancy effects when the experimenter and the subject were of the same sex ($z=+1.80$). But opposite results were found by Adair (1968) using a numerosity estimation task ($z=2.33$) and by Silverman (1968) using a reaction time measure ($z=1.61$). Both experiments found greater expectancy effects when experimenters and subjects were of opposite sexes. The best that can be said is that the sex of the experimenter does play a role in determining experimental outcomes, but its effect interacts in some unclear way with the sex of the subject and the nature of the experimental task.

Data on psycho-social characteristics which distinguish experimenters who show greater vs. lesser expectancy effects have been obtained by having subjects rate their experimenter's behavior on various behavioral items and by having independent judges rate the experimenters behavior from a movie of the E-S interaction. Experimenters exerting greater expectancy effects were found to be higher in: (1) Professional stance: acted in a more important manner, professional, businesslike, and consistent in behavior. (2) Inter-personal style: acted in a more relaxed manner, interested, enthusiastic, and personal. However, if the experimenter was overly personal in manner and made the experiment a "social hour," hence losing the "professional status," the effects of experimenter expectancy were reversed. (3) Kinesic communication: use of subtle kinesic signals from

the leg and head regions. However, if the movement patterns became too overt and hence detracted from the experimenter's "professional status," his effectiveness as an influence agent tended to diminish. (4) Paralinguistic communication: spoke slowly and in an expressive, nonmonotonous tone.

It would seem that these four dimensions can be assimilated under a "professional status" dimension.

In sum, Rosenthal (1966) notes that "the overall impression we have of the behavior of the experimenter who shows greater expectancy effects is that he is professional, competent, likable, and relaxed, particularly in his movement patterns, while avoiding an overly personal tone of voice that might interfere with the business at hand." This conclusion has received further support in research by Bootzin (1970) who found that more dominant experimenters obtained greater expectancy effects.

III. Evidence from Field Research

Main Effect.-- More recently, Rosenthal and other investigators have moved beyond the laboratory to the natural situation of the classroom in an attempt to assess the effects of teachers' expectations concerning school-related performance. To date thirteen studies have been reported dealing with the student-teacher dyad.[1]

---

[1] Of the thirteen studies, only four have been published. For the balance, I have relied on a summary prepared by Rosenthal (1969). See Table 2, Appendix A, for tabulation of the thirteen studies with their associated *z* values.

In the widely reported study by Rosenthal and Jacobson (1968),[2] all of the children in an elementary school in a lower socio-economic school district in California were administered a non-verbal test of intelligence. There were eighteen classrooms--three at each grade level and within each class 20% of the children were randomly assigned to the experimental group. The teachers were led to believe that the test was a predictor of intellectual "blooming." They were given a list of the children in the experimental group and told that great gains in intellectual competence would be shown by them in the following school term. The children were administered the post-test eighth months later at the end of the school year. The pre-and-post tests were actually Flanagan's Test of General Ability (TOGA which was chosen because it was "designed to provide measures of basic learning ability" (Flanagan, 1960, p. 6). TOGA may be viewed as having two sub-tests, one of reasoning ability and the other of verbal ability.[3]

Results indicated that for the school as a whole, the experimental group showed a gain of 3.80 points on TOGA, over their controls ($p < .02$). For the verbal subsection of TOGA the gain for the experimental children was only 2.06,while for the reasoning subsection it was 7.13.

A sex difference emerged in which the girls showed greater gains in comparison to the boys. For the total TOGA scores, the

---

[2] In the following section, several critiques of this study will be summarized.

[3] A critique of TOGA's appropriateness as a measure will be discussed later.

girls gained 4.7 points while the boys gained only 2.9. But in the reasoning sub-section, the girls gained 17.9 points while the boys actually lost -3.9 points ($p < .0002$).

Children in the lower grades tended to benefit more by teacher expectency than those in the upper grades. The first grade children in the experimental group gained 15.4 points over their controls while the second graders gained 9.5 points over their controls. However, in grade three, the experimental and control children showed no difference; in contrast, in grade four, the "bloomers" actually were slightly behind the controls in total TOGA score gains.

This study was replicated by Rosenthal and Evans (1968), this time with a middle class population in a small Midwestern town. Again, a significant main effect was found but now it was the boys who showed the greater gain in TOGA scores relative to the girls.

Following these two studies, Anderson and Rosenthal (1968) extended the work to retarded children. The subjects were twenty-eight boys attending a summer day camp eleven were randomly designated as "bloomers." In addition, six boys were chosen from the experimental group to participate in a program of special tutoring. Negative results were obtained, in that the boys who were exposed to the tutoring relationship as well as being designated as "bloomers" showed a decrease in reasoning TOGA scores, over the other groups ($p < .03$). For the "bloomers" in the untutored group, a non-significant expectancy advantage of just over 2 points was shown.

Follow-up testing was conducted several months after the end of the basic experiment. At that time, the tutoring "bloomers" had made up the expectancy disadvantage they had shown just after two months. The boys who received either the tutoring or the favorable expectation showed a 10 and 12 point expectancy advantage respectively, over their controls. Rosenthal (1968) interprets these findings as indicating that when these boys had both treatments simultaneously (tutoring and favorable expectation), they may have felt too much pressure. As a result, the treatments were ineffective.

Conn, Edwards, Rosenthal and Crowe (1968), working in a suburban upper middle class school, randomly designated 23% of the children in six elementary school grades as "bloomers." In addition to TOGA, these "bloomers" were also administered a task designed to measure the accuracy of their perception of vocal communication by male and female speakers. Results indicated that the children designated as "bloomers" tended to gain more in total TOGA scores than did their controls; but the difference was non-significant.

It was also found that the greatest benefits of favorable teacher expectations accrued to those children who were more accurate in their judging of the emotional tone expressed in adult female voices. Elsewhere (Rosenthal, 1966), it has been reported that auditory cues are important in the mediation of expectancy effects. The results of this study would tend to support that conclusion.

Studies of expectancy effects continued with research by Clairborn (1966) and Rosenthal and Anderson (1969). Clairborn, working with first graders, and Rosenthal and Anderson with older children, both found a tendency for children they designated as potential "bloomers" to gain less in TOGA scores than did the children of the control groups. Likewise, Pitt (1956) found no effect on achievement scores of arbitrarily adding or subtracting ten I.Q. points on the children's records. And Heiserman (1967) found no effect of teacher expectations on her seventh graders' stated levels of occupational aspirations. On the other hand, Biegen (1968) and Flowers (1968) both reported positive findings. In these two studies the entire classroom composed the experimental or control group, rather than specific children within a classroom.

A study by Burnham (1968) used s imming ability as the major dependent variable. His subjects were a group of pre-adolescent boys and girls attending summer camp for the disadvantaged. None of the children could swim when they arrived. A randomly chosen group of them was designated as the "high potential" group to the counsellors. On the presumed basis of a battery of psychological tests, the children so designated showed greater improvement in swimming ability than did their controls.

Beez (1968) reports a study of sixty Head Start children enrolled in a summer program. Of them, fifteen boys and fifteen girls were randomly assigned to the "low ability" (i.e., teacher expectancy of poor symbol learning) and to the "high ability" (i.e., teacher expectancy of good symbol learning) groups.

Teachers were asked to teach the children a series of symbols and jigsaw puzzles during a ten minute period. Results indicated that teachers who had a favorable expectation tried to teach more symbols than did those who had unfavorable expectations ($p < .001$). In addition, the children in the favorable expectation group learned significantly more symbols than did those in the unfavorable expectation group (5.9 vs. 3.1, $p < .001$) Rosenthal (1969) notes that Beez's procedure gives a better estimate of expectancy effects as they exist in the real world of the classroom where strong negative expectations doubtless exist- a point to which we shall return later.

Finally, Meichenbaum et al. (1968) studied a group of fourteen adolescent girls who were institutionalized at a school for offenders. The main variable under investigation was appropriateness of classroom behavior. Prior to and during the entire experiment, observations were made of the behavior of both teachers and students. The six girls chosen at random as "potential bloomers" showed a significantly greater improvement in classroom behavior than did the children in the control group ($p < .02$).

Rosenthal (1969) notes that of the thirteen studies dealing with teacher expectancy, there were five that showed no significant main effect. However, of these, three showed significant interactions of teacher expectation with some other primary variable (special tutoring, sex of pupil, etc.). Rosenthal concluded that the overall p value of the main effect of teacher expectancy is less than 1 in 5000.

Critique of School Studies

There have been several reviews of the 1968 work by Rosenthal and Jacobson. Snow (1969) and Thorndike (1968) feel that the use of TOGA was inappropriate since it does not have norms for the lower elementary grades and those from lower socio-economic backgrounds. Some wild intelligence scores are reported For example, for grade one, a mean score of 58 is reported--a highly unlikely situation. Furthermore, the published conversion tables do not go below an I.Q. of 60 and hence the tables had to be extrapolated. Some change scores were erratic. One subject, for example, had a pre-test reasoning intelligence score of 17 and post-test scores of 148, 110 and 112. It will be recalled that the main effect was demonstrated most clearly for the lower grades and for the reasoning sub-section of TOGA--the very areas criticized. Twenty per cent of the initially tested subjects were lost to the later testings and the effect of this loss on various subgroups is not dealt with.

Gumpert and Gumpert (1968) on the other hand, take note of the weakness of the normative TOGA data for the lower grades but feel that this does not invalidate the study but rather simply means that the results should be interpreted with "caution". They re-analyzed Rosenthal's data taking the classroom average as the unit of analysis rather than the individual child and found that the basic conclusions held up.

Rosenthal (1969) notes in his rebuttal that the reasoning subsection of TOGA did in fact show up significantly at some higher grade levels and not just at the lower grades as asserted

by the critics. Even if the three first grade classrooms are eliminated from the analysis, there still remain thirteen of fourteen classrooms that showed the hypothesized outcome ($p < .0006$). In short, the results do not depend on the inclusion of the particular classrooms singled out by Snow and Thorndike for criticism.

Rosenthal then takes up the general question of the validity of TOGA. The kindergarten teachers administered the pre-test to their children and, without knowledge of the TOGA scores, assigned these children to the fast, medium, or slow tracts of the first grade. TOGA was able to predict significantly which tract the children would be assigned to. Other evidence is presented to lead Rosenthal to the assertion that within the study there is sufficient evidence to defend TOGA as an adequate measure of total as well as reasoning I.Q.

Mediating variables.--Gumpert and Gumpert (1968) reflect the sentiments of many researchers in this area when they state that "virtually nothing is known about how they (expectancy effects) are communicated." The research that is available on mediating variables in field research will be summarized below. (See Table III, Appendix A.)

Little evidence was found by Rosenthal and Jacobson (1968) to support a differential time theory. Indirect evidence from this study suggests that the control and experimental groups were given equal time allocations by the classroom teacher. The theory of robbing time from control group subjects in favor of experimental group subjects suggests a negative correlation

between the gain scores made by the experimental and control groups within each classroom. In fact, a significant positive correlation (r=+.58; $p < .02$) was obtained.

More direct evidence bearing on the differential time theory came from a post-test questionnaire administered to the teachers. They were presented with a list of four children, two control and two experimental, matched by sex and TOGA score and asked to estimate the time spent with each child. Findings indicated that none of the obtained mean differences were significantly different from zero.

Although the Rosenthal and Jacobson study (1968) did not lend itself to an analysis of interaction quality, Rosenthal (1966) concluded from a summary of studies on the warmness variable that "warmer experimenters tend(ed) to obtain more competent and more pleasant responses from their subjects."

Weiner (1969) has recently offered an interpretation of Rosenthal's results. He proposes an attribution (cognitive) model of motivation in which ability, effort, task difficulty and luck are the major components. They can be classified as either stable or unstable as either internal or external sources of control. Weiner notes that the false information given to the teachers conveyed that the selected student had special abilities. Hence, occasional poor performance among this group could reasonably be attributed to a lack of effort. Attributing failure to deficiencies in effort has several consequences. Weiner's data revealed that one consequence is that "high ability children are castigated more for failure than "low ability" children. His suggestion is that differential punishment

indirectly leads to enhanced intellectual growth. That is to say, the experimental "high ability" group's occasional failures were attributed to a lack of effort rather than to lack of ability and hence were not punished. It is conceivable that informing children that their failures were not due to low ability will indirectly augment later performance. Attributing failure to a lack of ability implies that success is not possible in the future, while attributions to insufficient effort intimate that instrumental action can be undertaken which will lead to goal attainment. Weiner's main point is that conveying false ability expectancies have implications for teacher attributions concerning the causes of success and failure, and that these attributions conceivably can influence subsequent performance.

Earlier, the design of the study by Meichenbaum et al. (1969) was outlined in which a significant main effect was found. The main dependent variable under investigation was appropriateness of classroom behavior. Observers indicated whether behavior was appropriate or inappropriate and rated it inappropriateness on a three point scale for level of intensity. Teachers' behavior was also rated as positive (i.e., encouraging), negative (i.e., castigation), or neutral (i.e., task oriented).

Four measures were used to describe the frequency and nature of teacher-student interactions. The first was the percentage of all (%-ALL) teacher-student interactions which was directed toward each girl. This was composed of three

subcategories, namely, percentages positive (%-POS), negative (%-NEG) and neutral of teacher-student interactions. In sum, the %-ALL score reflects the amount of attention or teacher contacts the girls in the expectancy and control groups received. The %-POS and %-NEG scores indicate the relative frequency of teacher contacts differing in affective tone.

Results indicated that two teachers significantly increased their positive interactions with the expectancy subjects. Another significantly decreased her %-NEG interactions for the expectancy subjects, while the last teacher decreased significantly her %-POS interactions with the experimental subjects. Expectancy effects were not mediated by teacher's increased attention (%-ALL) to the expectancy subjects, but rather to the changes in the quality of interaction with subjects in the expectancy group. These results are consistent with Rosenthal and Jacobson's (1968) findings that a differential time theory could not account for the mediated higher performance levels.

Brief mention should be made here of the other school studies that contributed in part to knowledge concerning variables that mediate expectancy effects. Conn, Edwards, Rosenthal and Crowe (1968) found that expectancy effects were greatest for those subjects who were most accurate in judging the emotional tone of female experimenters' voices. This finding is consistent with results from laboratory research where it was found that the auditory channel was critical to

the communication of expectancy effects. Finally, Beez (1968) found that more material was taught to the expectancy group than to the control group.

These four studies represent the accumulated field research through 1969 on the mediation of expectancy effects. The results would seem to indicate that a search for mediating factors should be directed toward interaction quality rather than gross frequency.

IV. Recent Studies - 1969 and 1970

Recent studies are treated here separately because generally they employ a more complex and elegant design in which parameters have been systematically varied. In particular, task, criterion, characteristics of E and S, type of independent variables and theoretical stance have been examined. (See Table IV, Appendix A.)

The role of experimenter bias in verbal conditioning recently has received attention. Kessell and Barber (1968) reviewed 53 investigations of the experimenter-subject interaction variable in verbal conditioning from an expectancy point of view. The theoretical framework of the research was in terms of the construction of interpersonal attraction and interpersonal influence. The authors concluded that "it can be generalized...that the E-S relationship is a powerful variable in verbal conditioning." However, they admit that the effect is complex and not always predictable.

Kennedy (1969), continuing work in the area of verbal conditioning, varied expectations of E (positive, neutral, and

negative) as well as mode of stimulus presentation (auditory, facial, or gestural). Negative results are reported for both E expectation level and mode of presentation.

Sheehan's (1969) research has helped clarify the conflicting findings in this area. He postulated that the role of awareness was a confounding factor and led different groups of experimenters to expect that verbal conditioning either could or could not occur without awareness. Results indicated that subjects who were unaware of the correct contingency conditioned only in the group tested by experimenters who were led to expect that result.

These three experiments in the area of verbal conditioning are offered as an example of the direction experimenter expectancy research is taking--toward a refinement of parameters surrounding the effect.

A similar direction has been noted in recent school studies. Brager (1969) worked with fourth graders in six elementary schools using three treatment conditions: positive, neutral, and negative bias. Three tasks were chosen which varied along the dimension of amount of required E-S interaction. Results were significant for the tasks requiring maximum E-S interaction and opportunity for E-S interaction facilitated the experimenter bias effect. Kester (1969) employed a design using seventh grade pupils in six junior high schools testing a complex of ten hypotheses. Students were pretested within the first 2 days of school in mathematics, English, intelligence and school related attitudes. Teachers were assigned an

expectancy and control group and predictions were made in four areas of student functioning: pupil performance on achievement type tests; the quality of teacher-student interaction; pupils' attitudes as related to teacher expectations; and positive teacher communication as related to positive pupil communication. Observations were made of teacher-student interactions over the following nine weeks at which time post-testing occurred.

Of the ten hypotheses, four were significantly supported, three others were partially supported and three were not statistically significant. Kester did find that teachers communicated with their allegedly bright pupils in a more friendly encouraging, accepting manner as well as spending more time communicating with them. There was also evidence that as the pupils' positive communication to the teacher increased, the teacher's communications to the pupils tended to be shaped positively. There was no effect in terms of change in intelligence scores, attitudes toward school related subjects, or in achievement test scores.

Haskett (1969) pre-tested 267 Mentally retarded children on the Syracuse Scale of Social Relations (SSSR) and several other variables. SSSR scores for half the class were adjusted up or down and reported to the teachers. Teachers predicted the academic and social development progress of their students, as well as for a hypothetical class of children containing over, under, and average achievers. Post-testing occurred five months later. Videotapes were taken as well, as a measure

of time devoted to individual pupils.

Although the results are complex, they indicated that teacher expectations were related to pupils' social development, to academic achievement, and to several other variables. Haskett notes that "pupil performance tended to follow the direction of teacher expectancy even when the expectancy was not based on fact."

Finally, a series of experiments will be briefly described which highlight some of the problems in this area.

Barber et al. (1969) attempted to replicate the experimenter bias effect using the standard Rosenthal photo rating task. He added to the usual design a no-expectancy control group working with 51 experimenters and 501 subjects. Results were negative and he interpreted this as evidence for the weakness of the experimenter bias phenomenon.

Rosenthal (1969) in a rebuttal, first notes that the five experiments conducted by Barber et al. cannot be taken as serious replications of his work. There were differences in terms of experimenter status, sex of experimenter and type of subject. He then looked more closely at the data and observed a curvilinear relationship which is not sensitive to analysis of variance techniques. Rosenthal then re-analyzed the Barber et al. data, using chi-square statistics, and found a p value of .007 in support of the experimenter bias effect.

Barber (1969) in his rejoinder questions the appropriateness of chi-square, noting that it is arbitrary where the cutting point is made to produce a table. "Since the 501

subjects can be partitioned in many other ways, many additional chi-square tests can be performed. Of course, if we continue to carry out postmortem statistical anaylsis we may expect by chance to find some significant results." For example, Barber used the median as the cutting point and found negative results. Another cutting point was then chosen which produced positive results.

Levy (1969), commenting on the above dialogue, finds the choice posed by their conflicting research findings (i.e., either Barber et al. did nor replicate the Rosenthal paradigm or they did and the experimenter bias effect is thus disproved) to be an impossible one. The choice requires either a theory which states the parameters involved in the phenomenon or a body of systematic research from which the parameters could be induced. Levy finds both conditions lacking in this research area and concludes that it is a "moot question" whether Barber's findings contradicted those of Rosenthal. He suggests that the experimenter bias effect could be looked at in two ways: (1) as a methodological problem--a "spoiler effect;" or (2) as a member of a larger class of phenomena such as interpersonal influence or the effects of expectancies and values on perception and behavior. In either case, a taxonomy of situations and tasks is needed. He calls for a shift of emphasis away from the experimenter bias effect toward more fundamental methodological and substantive concerns.

This is exactly what has been attempted in several recent experiments. A study by McFall et al. (1970) will be described

in detail as representative of this type of work. Studies by Jacob (1970) and Rosenthal (1969) are along similar lines but will not be discussed here.

McFall et al. (1970) began with a communication model of expectancy effects and placed their study within the larger conceptual area of dyadic unintentional communication. They wanted to study subjects' characteristics which are related to experimenter expectancy bias. Thus they controlled one side of the dyad--the experimenter--by having tape recordings of E's voice. In a pilot study they had isolated experimenters whose voice consistently produced various bias effects.

On the basis of a communication model of expectancy effects, they predicted that subjects high in need for achievement (N-Ach) would show the experimenter bias effect, whereas those low in N ach would not. High N-Ach personalities, according to McClelland et al. (1953) are preoccupied with issues of success-failure and hence are more perceptive of subtel E cues. They also predicted that field-dependent subjects but not field-independent ones would be significantly influenced by the experimenter bias effect. Witkin et al. (1954) has noted that field dependent people tend to be more susceptible to social influence.

A significant difference was found between the mean photo-ratings given by high N-Ach subjects to the two experimenter tapes; the direction of the difference was appropriate to the anticipated stimulus pull of the tapes. In contrast, no significant difference was found between the low N-Ach groups.

Likewise, subjects classified as field-dependent did show a significant difference between the photo-ratings they gave under the two experimenter instructional conditions. No significant difference was found in the ratings of field-independent subjects. There was also a tendency for subjects scoring high on both N-Ach and field dependence to show a larger expectancy influence effect than subjects scoring high on either variable alone. The authors conclude that the results enhance the status of the communication model of expectancy effects from which these hypotheses were derived.

Concluding Note

What are we to conclude regarding the reliability of the interpersonal expectancy phenomenon? Despite the statistical criticisms outlined above, the trend of the findings generally supports the assertion that interpersonal expectancy effects are a reliable and measurable phenomenon. In a massive study of educational opportunity, Coleman (1966) supports this assertion by his finding that "teacher characteristics" accounted for a large part of the between school variance. Although he did not directly investigate teacher expectancy as such, this variable is likely to be one of the components of "teacher characteristics."

There is a widespread belief in the educational community that negative teacher expectations account, in part, for Negro underachievement. Clark (1965) has asserted that "(Negro) children who are treated as if they are unteachable invariably

become unteachable." Epps (1969) has edited a review of recent findings on Negro underachievement. Although teacher expectation was not investigated as such, there is tangential evidence suggesting its importance in producing Negro underachievement.

I am in agreement with Gumpert and Gumpert (1968) who concluded that interpersonal expectations "appear to be quite strong and stable in spite of the subtlety and simplicity of the experimental induction." What is called for now is further research on "how" teacher's expectations of children become realities.

V. Theoretical model and predictions

One of the models of expectancy mediation considered earlier was that of an operant reinforcement paradigm. The purpose of the research planned here is to test this model as a mediating mechanism in expectancy effects. Simply stated, such a model predicts that behavior can be controlled by the powerful effects of appetitive and aversive stimuli. That is to say, the proper temporal presentation of stimuli possessing positive reinforcing qualities increases the probability that selected behavior will increase in frequency. Conversely, stimuli possessing negative reinforcing qualities will decrease the probability that the selected behavior will re-appear (Skinner, 1953). Honig (1966) has shown how this model has produced behavior modification in diverse areas of inquiry.

Canavan (1968) and Gumpert and Gumpert (1968) have offered

a descriptive account of how an operant reinforcement model would operate in the classroom. They ask us to imagine an unremarkable appearing boy who is chosen as one who is about to experience a period of intellectual blossoming. Given a situation in which the teacher is trying to communicate a bit of information, she is now more likely to expect an indication of comprehension than if she had been told nothing about potential academic progress. A minor change in his behavior at this point, say a nod or smile, may be interpreted as a glimmer of understanding so that, encouraged, the teacher intensifies her efforts to reach him. Consequent subtle changes in the teacher's behavior and attitudes, such as alterations in her body posture, tone of voice, perceived interest, facial expression, or verbal praise, may similarly interest and encourage him, leading to increased motivation and attention, and finally to the reward of mastering something new--as well as the fulfillment of his teacher's expectation. In this way, the teacher's and student's behavior are shaped through mutual reinforcements.

Homme (1963, 1965) has shown with his research on "contingency management" how a positive reinforcement paradigm operates in the classroom. He adopted Premack's P-Hypothesis or more commonly called "the differential probability hypothesis" which states that "any higher probability behavior will reinforce any lower probability behavior upon which it is contingent." Setting out to control the behavior of three year olds without the use of punishment or candy-trinkets as

reinforcers, Homme and his associates identified high probability reinforcing behavior in the children's own terms: running and screaming! They made engaging in these behaviors contingent on low probability behavior: doing a small amount of "sitting quietly in chairs and looking at the blackboard." Homme claims that through this method they were able to teach in about one month all the material usually taught in first grade. Other cases are presented by him to illustrate the usefulness of a reinforcement paradigm in the classroom.

It will be recalled that within several limited laboratory experiments, a reinforcement model was not supported (Rosenthal, 1966; Masling, 1965; Weick, 1965; Adair & Epstein, 1967). The essential requirement was to demonstrate that responses occuring later in the experimental session would be more in line with E's expectation than earlier emitted responses. The results from all four of the studies bearing on this proposition were non-supportive. However, the test of the operant model was limited to single, time restricted laboratory experiments and did not consider the reciprocal reinforcements occuring between experimenter and subject.

A broader conception of the reinforcement model as a mediating mechanism was tested by examining a series of E-S interactions over a series of experimental sessions. The prediction was made that later seen subjects would show the desired effect more clearly than earlier seen subjects. In all, 13 studies (see Table 1, Appendix A) bear on this "interpersonal learning" hypothesis. Seven supported the

hypothesis while six yielded non-significant results.

In contrast to the equivocal results in laboratory settings, Meichenbaum's (1968) classroom experiment points to a positive reinforcement model as a mediating mechanism. However, his study failed to specify the specific nature of the positive interaction; nor were specific predictions made concerning the expected results. Brager (1969), on the other hand, found that on the basis of positive, negative and neutral feedback, experimental groups could be differentiated on the dependent variable.

The primary purpose of the current research is to test the applicability of a reinforcement paradigm in the mediation of expectancy effects. This will be accomplished in three ways:

(1) Experimental Group Analysis: Creating four experimental groups with different levels of reinforcement possibilities.

(2) Item Progression Analysis: Gathering data on the sequential aspects of responding on the dependent variable by analyzing blocks of responses.

(3) Mode of Reinforcement and Interactive Analysis: Making tape recordings during the post test to permit the measurement of changes in voice tone and quality, as possible reinforcing stimuli. The recordings will also provide information on the interaction between teacher and student during the post-testing.

Positive and negative expectations will be induced in teachers and students in various combinations. In all, four

experimental groups will be created. For two of them the induced expectation will be congruent for teacher and student (i.e., both student and teacher will receive identical expectations); whereas for the other two they will be incongruent (i.e., teacher will receive positive and student negative expectation). In addition, an average expectancy control group will be created.

Schematically, the design appears as follows:

		STUDENTS		
		NEGATIVE	POSITIVE	AVERAGE
	NEGATIVE	GP. 1	GP. 4	--
TEACHERS	POSITIVE	GP. 3	GP. 2	--
	AVERAGE	--	--	GP. 5*

Predictions based on an operant reinforcement model call for Groups 1 to 4 (experimental groups) to show shifts on the dependent variable relative to Group 5 (control group) as follows: Groups 1 and 4 to shift downward on the dependent variable and Groups 2 and 3 to shift upward on the dependent variable. In addition, Group 2 will show maximum upward shift and Group 1 maximum downward shift on the dependent variable in comparison to Groups 3 and 4. Group 3 will shift upward

---

* A word of explanation is in order regarding the control-average expectancy group. One choice was to create a no-expectancy control group. For several reasons, an average expectancy group seemed more appropriate: 1) The absence of expectation was a highly unlikely situation. If the control group was told nothing it is likely that they would have created their own idiosyncratic expectations yielding a heterogenous control group which would overlap with the experimentally induced ones. These idiosyncratic pre-experimental expectations again would yield a heterogenous control group.

on the dependent variable and fall between Groups 2 and 5; while Group 4 will shift downward on the dependent variable and fall between Groups 1 and 5.

The rationale for these predictions arises directly from the operant reinforcement model of behavior control. The congruent positive group (no. 2) contains the maximum possibility for showing upward shifts on the dependent variable. With both teachers and students expecting positive student performance, the complex of covert cues and signals between teacher and student (body posture, voice tone, facial expression, etc.) should reinforce each other's positive expectation and produce maximum upward movement on the dependent variable.

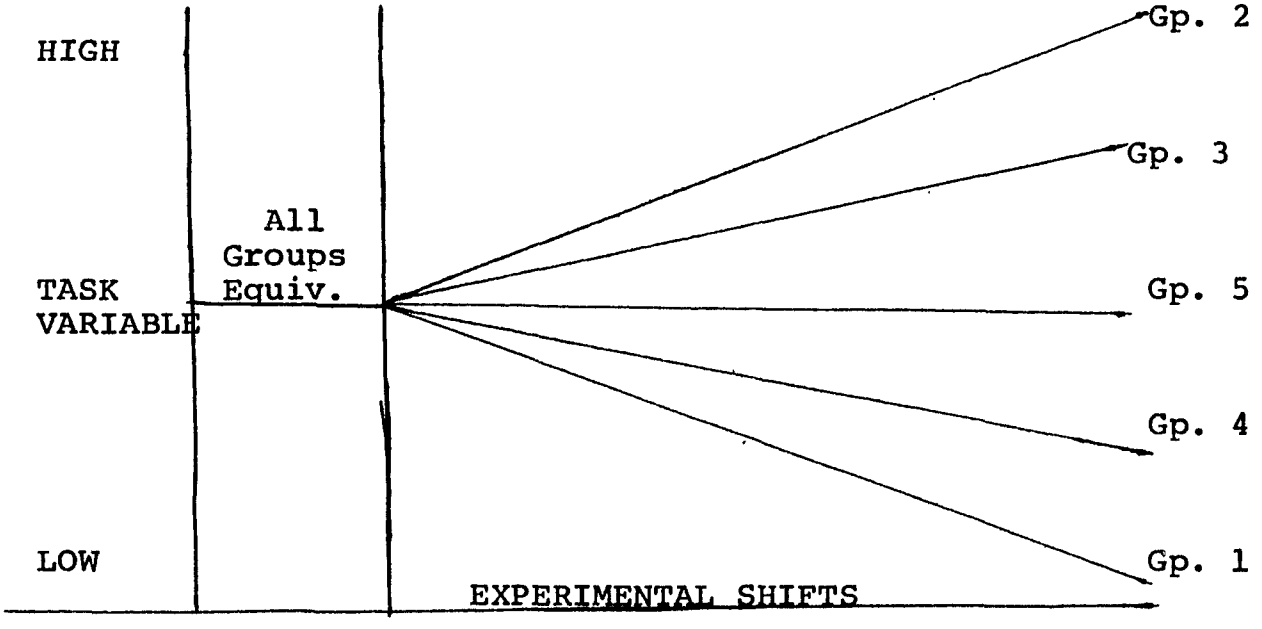
Conversely, the congruent negative group (No. 1) will contain the maximum possibility for showing downward shifts on the dependent variable. With both teachers and students expecting negative student performance, the complex of covert cues and signals between teacher and student should reinforce each others' negative expectation and produce maximum downward shifts on the dependent variable.

The incongruent expectation groups (Nos. 3 and 4) should fall somewhere between these extremes. Reinforcement theory would suggest that teachers are more important than students as reinforcing agents since they provide reinforcers of greater frequency, subtlety and significance. The role of "teacher" as well as "adult" contains powerful reinforcing properties in relation to students.

Group No. 3 (teacher positive and student negative) should shift upward on the dependent variable, since the reinforcing properties of the teacher should overwhelm those of the student. However, the upward shift should not be of magnitude as great as that predicted for the congruent positive expectation group, since both members of the dyad in the latter case are emitting positive cues.

The same reasoning applies to Group 4 (teacher negative and student positive). Since one member of the dyad is emitting negative cues, it was predicted that there would be downward shift on the dependent variable. But again the shift should not be as great as it is for the congruent negative group where both members of the dyad are given negative expectations. The teacher, being a more powerful reinforcing agent than the student, should overwhelm the positive cues emitted by the student.

The purpose of including the incongruent groups was to examine the effects of opposite teacher-student expectations as well as to provide a continuum for positive to negative expectations. Graphically, the predictions are as follows:



As the E-S interaction continues, the E and S should shape each other's behavior through differential reinforcement. Hence, responses later in the E-S interaction should be more in line with induced expectations than are earlier ones. To test for this sequential effect, blocks of items will be analyzed. For the congruent negative-negative expectation group (Gp. 1), reinforcement theory would predict that the average score for progressive blocks of items would decrease. Conversely, for the congruent positive-positive expectation group (Gp. 2), reinforcement theory would predict that the average for progressive blocks of items would increase in a gradual upward slpping line. The incongruent groups (Gps. 3 and 4) contain contrasting negative and positive reinforcement elements. For Group 3 (teacher positive, student negative), it was predicted that there would be an increase in responding as blocks of items progressed. However, since only one member of the dyad is emitting positive cues, the rate of increase should be less than when both members are. As noted earlier, reinforcement theory suggests that "teachers" are more powerful reinforcement agents than "students." Hence, their positive induced expectation should overwhelm the negative cues emitted by the students and a gradual upward slope of increase in responding should occur. Similarly, in Group 4 (teacher negative student positive), the teacher is emitting negative cues which, based on the above reasoning, should overwhelm those positive ones emitted by the student. Hence, for this group the average score for each block of items should decrease over

time. However, the decrease should not be as extreme as it is predicted to be for the congruent negative-negative group where both members of the dyad received negative expectations. It was also predicted that there would be no change over blocks of items for Group 5 (average-average), since reinforcement theory suggests that neutral feedback lacks the motivational impact of either positive or negative stimuli.

In sum, the prediction is that the five groups would rank order from high to low in terms of an item analysis on the post-test as follows: Groups 2, 3, 5, 4, 1. This pattern is identical to that formulated for the experimental group analysis.

Tape recordings will be taken during the post experimental sessions to permit analysis of the interactive aspects of response chaining between teachers and students and of the mode and nature of delivery of the reinforcing stimuli. Reinforcement theory suggests an interactive response chaining between teachers and students. Early in the E-S interaction cues are random and unprogrammed. As time passes, they become more stable as both members of the dyad work out patterns of mutually understandable cues.

The source of data for the students' patterning of responses is an analysis of their responses on the dependent variable (i.e., an item analysis as discussed above). For the teacher, the source of data will be changes in various aspects of vocal expression over the course of the post-testing session. Scheurer (1970) recently has developed several dimensions and scales on which vocal utterances can be rated.

Comparisons will be made between these two measures. Specifically, it is predicted that there will be parallel changes in the two measures over the course of a post-testing session. For Group 1 (negative-negative) both measures should decline (i.e., teachers giving off negative utterances and students progressively getting more items wrong) over the course of the post-testing session. However, for Group 2 (positive-positive) the measures should show continuous increases. Predictions for the other two experimental groups and control group are similar to those stated above.

There is sufficient evidence in the literature to indicate that the auditory channel is essential to the communication of expectancy effects. Reinforcement theory would predict that the teacher receiving positive expectation feedback would deliver more rewarding stimuli (i.e., speak in a "warmer" tone, etc.) than a teacher receiving negative feedback. The presence of such factors should be clearest in the congruous groups since both members are receiving similar feedback. In the dissonant groups, the student's conflicting feedback presumably could act to suppress at least partly the reinforcement being emitted for the teacher. However, as noted earlier, reinforcement theory would predict that teachers are more powerful reinforcing agents than students. Hence, the effects of their reinforcement should not be overwhelmed by the students but only partially suppressed. Predictions are identical to those outlined in the two earlier sections.

Based on a review of the literature, there were several goals which were delineated in designing this research. One was that mediating variables as well as data bearing on the interaction between teacher and student should become the focus of the research. These points have already been discussed. It was also desired that the success of the experimental induction be established separate from the effects of the dependent variable. In the following section this will be elaborated on and delineated. Finally, it was desired that the research be conducted in as naturalistic a setting as possible. Argyris (1969) recently has noted that a major shortcoming of social-psychological theory is the restriction of its findings to laboratory settings. The research planned here will be conducted in a school setting.

VI. Method

Definition and Measurement of Independent and Dependent Variables.

The major independent variable under investigation was expectancy. The following considerations influenced the choice of the specific behavioral area in which the expectancy was created: (1) the area should not be too vague, nor so specific that it would lack generalizability; (2) the behavior should be open to change as a result of experimental manipulation, but be reasonably stable or reliable under ordinary circumstances; (3) the area should be one where the teachers had little or no prior testing experience, in order to make the experimental induction creditable to them; and (4) the area chosen should be one that is not relatively

critical for childhood development.

Three classes of measures were conceptualized: (1) those that are expected to be directly responsive to the independent variable; (2) those that may reveal generalized effects of the independent variable on non-related task behavior; and (3) those that deal with some socio-emotional attitude. Rosenthal (personal communication, 1969) has suggested that the latter is a way of determining whether the experimental manipulation has any "spill-over effect" into a non-task area of behavior.

It should be made clear that the experimental manipulation and the associated predictions apply only to the first type of measure. The two other types, which were not the focus of an experimental induction, are included here in order to gain further data on the generalizability of the effect.

On the basis of the above considerations, the following measures were chosen:

1) Iowa Test of Basic Skills (Class One Measure).
2) Coding sub-test of the Wechsler Intelligence Scale for Children (WISC: Class Two measure).
3) Self-Esteem Inventory (Class Three measure).

Iowa Test of Basic Skills: This is a test developed at the University of Iowa under the direction of E.F. Lindquist and his colleagues. It has a long history of development and refinement, has been widely used in elementary school systems throughout the country and has received a favorable review in Buros (1961). National percentile norms are available as well as special norms for Catholics, large

cities, the Northeast, Southeast, Midwest, Southwest and Far West schools. Although it contains 15 sub-tests, the one chosen for use in this study is called "Knowledge and Use of Reference Materials" (W-3).

The test catalogue states that this sub-test is concerned wholly with the fundamentals of elementary school instruction (i.e., with the basic skills essential to success in any type of schoolwork). Since the test battery measures the pupil's ability to put to use his acquired skills, no test or sub-test is concerned with repetition or identification of formal facts or rules. Tests for each grade are adapted specifically to that grade, with the addition of some test items which are appropriate for measuring the extreme ranges in the grades tested. The test contains four equivalent forms for each grade level.

It should be noted that for the fourth grades used in this study, the third grade version of the Iowa was used. The children actually were at a third grade level of functioning, since the school term had just begun and they were only a few weeks into fourth grade materials. Similarly, the fourth grade version was used with the one fifth grade.

As developed by the test publishers, the fourth grade version contained 42 items and the fifth grade version, 52 items. See Appendix B for the forms of the tests used as the pre- and post-measures.

The test was modified in two ways to suit the purposes of this study: (1) The index of item difficulty was equated

for the first and last half of the test. As it stood, items became increasingly difficult as the test progressed. Since one aspect of the data analysis called for an item analysis to examine sequential effects, this confounding of item difficulty with order had to be eliminated. Hence, items were rearranged so that the first and second halves of the test were equivalent in difficulty. (2) Ten items were added to the test, since, with item difficulty rearranged, there was a series of difficult items early in the test. It was felt that this might discourage students early in their test-taking and hence have an adverse effect on performance. Consequently, ten items which were one grade below the students' current grade level were added to the test. Seven of these additional items were the very first ones in the test while the remaining three were added at the end. As a result, the test for the fourth graders contained 52 items (42 regular Iowa items plus 10 additional).

Reliability and validity data are reported in the Houghton-Mifflin Test catalogue (1964). For the third grade version, the test was standardized on a sample of 18,496 students throughout the United States. A sub-sample of 12.5% were chosen for a split-half reliability anaylsis yielding a coefficient of .87 for the W-3 sub-test. For the fourth grade version, a similar procedure was used yielding a reliability coefficient of .88 for the W-3 sub-test.

In terms of predictive validity, the authors note that the test was "not designed as an aptitude test nor as a

predictor of future academic success. However, the importance of proficiency in the basic skills in determining high school and college success has been demonstrated repeatedly."

They continue by noting that "evidence of true predictive power of the tests is difficult to obtain because the selective process of education tends to eliminate from research samples a large proportion of pupils whose low level of performance could be predicted with a high degree of certainty." Nevertheless, several studies of predictive validity have been completed. One was a 1958 study comparing Grades 4, 6 and 8 Iowa scores with Grade 12 scores on the Iowa Tests of Educational Development. The zero-order correlations were .81, .82, and .77 for Grades 8, 6, and 4 respectively. The authors present other studies that found the same magnitude of relationships as reported above.

The rationale for the selection of this test as the main dependent measure is that it meets the requirements outlined earlier. The area chosen--"Work-Study Skills"--is not so vague as to be meaningless and likewise it is not so specific as to lack generalizability. Work-study skills are essential for performance in many academic areas. In effect, this measure cuts across specific areas of functioning and abstracts a skill needed for adequate performance in all of them. In addition, it is thought that the test taps an area of functioning of specific students which the teachers will have few pre-conceptions about. Testing occurred just two weeks after school began in September, 1970, so that the

teachers had little experience with the work-study skills the students possessed and probably would not have strong convictions along these lines. Upon entering the school, I learned that all the third grade teachers from the previous year had resigned from the school. Hence, the current fourth grade teachers used in this research could not refer to them to gather information about the children. Their only source of pre-conceptions were the students' permanent school files which they customarily perused as school began. It was also thought that work-study skills was an area of functioning with which the teachers would have had little prior experience regarding test administration. Reading, mathematics, science, spelling and group intelligence tests are routinely administered in classrooms over the years, but not a test of work-study skills involving the use of a dictionary and index. Consultation with the school principal revealed that the Iowa test had never been administered in his school and that in general the school was not test oriented.

In addition, the area of work-study skills seems less essential to childhood development than such things as, for example, reading. This, of course, is a relative matter since all are of some importance in developing the whole child.

Coding Subtest of WISC: This is a two minute test (see Appendix C) in which the subject is asked to transfer codes from a key onto an answer sheet. Wechsler considers it a supplementary test, to be added to the total WISC if time permits. It is considered here as a measure of "effort" or

motivation, since a student who has been given a positive expectation might indeed try harder to fulfill that expectation in performance on this test.

Self-Esteem Inventory: Coopersmith (1967) recently has reported on a six year study of self-esteem using 8 and 10 year olds as his subjects. He begins by defining self-esteem as "the evaluation which the individual makes and customarily maintains with regard to himself: it expresses an attitude of approval or disapproval, and indicates the extent to which the individual believes himself to be capable, significant, successful and worthy." During the course of his studies, he developed a 58 item test of self-esteem which yields a "total" score plus five sub-scales. (see Appendix D) The "total" score is based on 50 items while the five sub-scales include fewer items as follows: (1) general self - 26 items; (2) social self--peers - 8 items; (3) home--parents - 8 items; (4) lie scale - 8 items; (5) school academic - 8 items. These scales follow in sequence through the inventory. Coopersmith notes one of the test's strengths is its ability to reflect changes in specific areas of functioning.

The test has been used with populations ranging from 9 years to adult level and there is no sex difference. In most samples, the curve is skewed in the direction of high self-esteem. He reports norms for pre-adolescent boys as 72.2 (based on a total possible score of 100) and 70.1 for girls.

The test-retest reliability obtained for the self-esteem inventory after a five week interval with a sample of 30

fifth-grade children was .88 and the reliability with a different sample of 56 children was .70. Coopersmith notes that in evaluating these figures it must be borne in mind that a child's picture of himself has just begun to become stable at this age level so that some of the apparent unreliability in the measure is due to maturational factors. Of those measures currently available, this measure would appear to be one of the most stable.

Pre-experimental Expectation Measures: The pre-experimental expectations of students (i.e., their own personal view of their ability on the Iowa test) and of teachers (i.e., their own personal view of each student's ability on the Iowa test) were measured at the pre-test time. This was necessary, since preconceptions on the part of students and/or teachers could interact with experimentally induced ones. It was felt that this was a factor that should be equalized (i.e., matched) across the five treatment groups.

Teachers' pre-experimental expectations were secured just prior to the pre-testing, on the ninth day of the school term. They were given a ten point scale (see Appendix E) called "Hunter Test of Study Habits Rating Form" ranging from 1 (poor) through 5 (average) to 10 (exceptional), and asked to indicate for each child in their class "where, in your opinion, the child falls on the scale." Teachers were allowed to indicate that they had "no opinion at this time."

As part of the pre-test, students were asked to indicate how well they thought they could "alphabetize and use

dictionaries and information books." A three point scale (see Appendix F) was used: 1 (good); 2 (fair); 3 (poor).

Tapes: Tape recordings were made of the post-experimental sessions. A cassett model number TC110 tape recorder was used with 60 minute tape cartridges. Prior to the post-testing, teachers and students were advised of the presence of the tape recorder, and told that it was necessary so that the E could study problem areas that arose during the testing. Initially the students were intrigued by the recorder; but they lost interest in it as soon as the post-testing began and they got into the rhythm of the testing. It was felt that the recorder's presence did not significantly alter the activities during the post-testing.

In all, 20 hours of tapes covering all post-testing sessions were secured. During the data analysis, they were used to identify differences in voice tone and quality among the teachers for the five experimental groups.

VII. Procedure

Equalizing Groups on Pretest Scores: Lindquist (1953) has noted that type "G"[X] errors are particularly prevalent in educational research where variations in teachers and classes are common. Hence, the students in each class were distributed over the experimental conditions, making the class and the associated teacher a factor in the analysis.

X Lindquist describes this as some property (i.e., better teacher, better classroom) that may favor one treatment at the expense of the others in the comparison of means.

Within each of the four classes, the five conditions were equalized on the following four attributes: sex, scores on the Iowa Test of Basic Skills, own pre-experimental expectations[xx], and teacher's pre-experimental expectations[xxx]. It was thought that these variables were critical ones since variations on them among the experimental groups could be related to differences on the main post-test measure. The groups were not equalized on scores on the self-esteem inventory or the coding test. This was due to the small N for each class. Analysis of covariance techniques were used to analyze changes on these measures.

Equalizing on verbal ability was not deemed necessary for two reasons: (1) the students were of middle class backgrounds in a suburban school system of considerable quality; hence, they appeared to possess fairly homogenous verbal ability. There were no non-English speaking children in this school system. (2) Items were read aloud by the classroom teacher. Any verbal difficulties encountered by children should be

---

xx Public relations entered in here. A few parents were known to be chronic complainers in the school district. Upon the request of the school principal, their children were included in groups given positive expectations so that if they discussed the experiment at home it would be within a favorable framework. This consideration never altered the mean score for any experimental group.

xxx The E was careful to include students in groups that confirmed teachers expectations. This was only for extreme ratings by the teachers and only for a few children. For example, in one case teacher 4 rated one child as a 1 (low expectation) and the E made sure to include him in a negative expectancy group. In no case did this consideration alter the mean score for any experimental group for any teacher. Including this child in a positive feedback experimental group might prove incredulous to the teacher and hence disturb her confidence in the experiment.

reduced by this procedure.

With these considerations in mind, the 100 students were assigned to the five experimental groups as outlined in Table I, Appendix M. Adequacy of equalizing the groups was tested by ANOVA. All of the relationships were non-significant as follows: $F=0.04$ for experimental groups on the Iowa; $F=0.22$ for experimental groups on teachers' pre-experimental expectations; $F=0.74$ for experimental groups on students' pre-experimental expectations. By inspection, the groups appeared to be equalized adequately by sex. A chi-square analysis yielded a non-significant relationship ($X^2=0.52$).

Execution of Experiment:--The execution of the experiment extended from May through November 1970 and is summarized in detail in Table II, Appendix M. Included here is a summary of the relevant considerations.

During the spring and early fall of 1970, a series of meetings was held with the school principal and Superintendent of Schools to elicit and soldify their support for the study. During the first week of the fall school term a meeting was held by the E with the four participating teachers and the principal to acquaint the teachers with the research. They were told that the main point of the experiment was to gain further validity and normative data on the "Hunter Test of Study Habits." (Actually this was the Iowa Test of Basic Skills). It was stated that the test had proven to be an excellent predictor of potential bloomers in the area of study habits. We now wanted to see how well it worked in a school district in this locale with this particular socio-

economic status.

It was stated that the first round of testing would be done in the classroom by a former, presently non-working school teacher. (This person was actually an accomplice of the E.) After the papers were scored, the E would return to the school for a second round of testing. This time the teachers would conduct the testing with small groups of students. It was stated that the teachers would then be advised of the results of the first testing. The teachers' pre-experimental expectation measure was distributed and returned before pretesting began.

Pretesting occurred the following week and was conducted by the accomplice with the aid of the classroom teacher. Post-testing extended over a ten day period and began some four weeks following the pretest. Teachers, experimental groups and order of post-test battery administration were counter-balanced. The pattern of post-testing proceeded as follows: The E and accomplice entered the scheduled class at the appointed time. The E accompanied the small group of students to be tested to another part of the building where all post-testing took place. This was a small music room especially equipped for the post-testing. The E read aloud the content of the "report form," which was the same for all the students in that group, and asked them to indicate whether or not they believed that that was true about themselves. The completed forms were collected by the E. Simultaneously, the accomplice was inducing the experimental induction with the teacher in a

similar manner (see Appendix L). Within a few minutes the teacher entered the room where the post-testing was to take place and immediately began. She read aloud all the test items and generally supervised the post-testing session. The E's role was to monitor the tape recorder; he sat quietly nearby. At no point did the students or teachers raise any questions about the validity of the experiment.

Some weeks later, the E met with the participating teachers to de-brief them. One teacher reported that she suspected the intent of the experiment but upon questioning, it appeared that she had not perceived its true intent. The other three had never had any suspicions. Actual test scores were distributed and the experiment discussed.

School and Subjects

The setting for this experiment was the Glen Head Elementary School located in Nassau County, New York. This community is located on the suburban North Shore of Long Island. It is a neighborhood school which generally has positive frequent contacts with the community. The students are drawn from parents of moderate income, although there are some families of considerable means. The parents' occupations are principally white collar. There are few children of lower class parents. In general, the school can be characterized as middle-middle to upper-middle class in its orientation and composition.

Subjects were 100 students in three fourth grade classes and one fifth grade class, as well as the four participating

teachers. The fourth grade was selected because (1) students' attitudes are still pliable at this age level and (2) the students are old enough to understand an explanation of the true intent of the experiment. There were no minority group or non-English speaking children in the classrooms tested.

The four teachers included one male and three females, three of whom had earned their Master's degrees in Education. Their average age was around 37 years and each had been teaching for from four to six years. The three female teachers had returned to teaching after raising their families. Both the teachers and students were considered typical of those found in suburban middle class communities in the New York area.

Some Ethical Considerations

Rosenthal's (1968) work was conducted only with positive and no-expectancy control groups. Negative expectations were avoided since it was felt that the long term effects of such negative expectations could possibly produce long term damage to the children. In contrast, the current study called for the creation of just such negative expectancy groups.

There were several points to be considered in justifying this decision: (1) Rosenthal's work dealt with global dependent variables (i.e., I.Q. scores, achievement scores) which may have an impact on the student's total view of himself. Here we were dealing with lower order variables which were more specific in nature and hence had less implications for long term damage. (2) Post-experimental sessions were held with both teachers and students to reveal the true intent of the

experiment. In this way the negative expectancy groups were told their actual ability. Grade 4 was selected for experimentation with the above considerations in mind. By that age children are generally capable of understanding a post-experimental explanation of the true intent of the experiment. Children of a younger age might not understand such an explanation and with them the impact of negative expectations might be damaging. (3) Beez's (1968) work is an example of research where negative expectations were created; and he found a dramatic effect of such expectations. Hence, it was felt the inclusion of the full comtinuum of expectations from positive through negative would make for a more elegant and complete design. In addition, the ecological validity of the experiment would be enhanced since both positive and negative expectancies exist in the real world of the classroom.

VIII. Results*

Iowa Test of Basic Skills

The first part of this section will deal with comparisons among treatment conditions. Table 1 presents the unadjusted means, standard deviations and change scores for both the Iowa pre and post tests and the ANCOVA adjusted post-test means based on the pre-test score as the co-variate. Table 2 indicates that the main effect of treatments was significant at the .05 level. The average-average group appears to have

* Analyses were conducted on the Harvard IBM 360 located at Cambridge, Massachusetts and the programs available at the Computing Center at The City University of N.Y. located at 33 West 42nd St., New York.

contributed principally to this finding being discrepant in a positive direction from the other groups. A main effect of sections was also found ($P = .01$). One of the four sections was discrepant from the others, although they appeared to not differ in any significant way.

In order to examine the difference between the five treatment conditions, the Scheffé test was applied to the ten available comparisons; none of the paired comparisons yielded a significant difference. However, the average-average group was found to be significantly different at the .01 level from the other four groups combined. On the assumption that the average-average treatment can be interpreted as a kind of control condition, Dunnett's method was used to compare the groups with each of the others. The average-average group was found to be different from the positive-positive group at the .05 level of significance and from the positive-negative and negative-negative groups at less than the .10 level of significance. Somewhat similar results were obtained when comparisons were made in terms of the net proportion who increased their Iowa scores (Hovland, Lumsdaine and Sheffield). The data are presented in Table 3. The average-average group was significantly different from the positive-positive and negative-negative groups ($P = .05$). Even though by sign test no group departed significantly at the .05 level from a null hypothesis of equal number as increase as decrease, the trend of the results presented in Tables 1 and 3 are opposite to the predictions. For example, the positive-positive group

was expected to show an improvement in performance on the Iowa test. Instead, the group as a whole declined in score ($P = .08$). Contrary to this, the negative-negative group was expected to decline in performance and instead the trend was toward improvement ($P = .16$). At the same time, the average-average group, for which no prediction was made, exhibited a general pattern of improvement ($P = .07$).

A multiple regression analysis was then carried out in an attempt to further understand the variance on the Iowa post-test. Analyses were completed for each of the five treatment groups, as well as for all subjects combined. Table 4 contains the relevant inter-correlation matrices for the entire sample. The predictor variates were: 1) Iowa pre-test score, 2) teachers' pre-experimental expectations, 3) students' pre-experimental expectations and 4) pre-test score on the self-esteem inventory. The best predictor of the Iowa post-test score was the pre-test score.

Beta weights were computed for each predictor variate for the total sample. For the Iowa pre-test the weight was .75; teachers' pre-experimental expectations, .02; for students' pre-experimental expectations -0.15; and for the pre-test score on the self-esteem inventory -0.03. Based on F tests, only the Iowa pre-test weight was significant ($P = .001$). This statistic was used to test whether the variable within the regression equation added any significant amount of predictive ability. The pre-test accounted for 62% of the observed variance on the post-test, with the other

three variates contributing nothing to the prediction provided by the Iowa pre-test. The analysis for the sub groups confirms the analysis for all subjects combined (See Tables 1 to 5, Appendix O).

Split-Half Analysis

The second part of this section deals with a split-half analysis of the Iowa post-test. Earlier it was predicted that items later in the response series would be more in line with induced expectations than ones earlier in the series. Tables 5 and 6 present the mean score on the Iowa post-test for the first and second sections of the test for each experimental group and the control group. Inspection of these Tables reveals very little change in the mean scores during the course of the post-test. Repeated measures ANOVA, based on the difference between the comparable halves of the pre and post tests, confirmed the absence of any significant main or interaction effects (See Tables 6 and 7, Appendix O). Hence, the prediction was not substantiated.

Tape Analysis

The third portion of this section deals with an analysis of tape recordings collected during the post-testing sessions. Twenty hours of tape recordings were collected covering all post-testing sessions. An analysis tape was prepared for rating by selecting 20 second samples of the teacher's voice as she administered the experimental task. For each post-testing session, three samples were taken: one at the very

beginning of the session, another in the middle and the last one at the very end. This procedure yielded 60 voice samples* (i.e., four teachers X five experimental groups X three samples per session = 60 voice samples). Following this, the tape was modified to conceal the content of the teachers' verbalizations and thereby remove it's influence on the raters. This was accomplished by feeding the tape through a "content filtering box" designed by Scheurer (1971) at the Harvard University Electronics shop. The result was a tape where 1) the content was sufficiently distorted so that it was impossible to understand what was said, 2) background noises (e.g., car outside the classroom, noises in the hall etc.) were eliminated and 3) the qualitative aspects of the voice (i.e., pitch, tone, quality) were heightened. The result is a voice which sounds like a person talking on another side of a wall.

Following the suggestions of Milnoe (1970) and Scheurer (1971), middle aged women were used as judges since it is reported that they have exceptional sensitivity to vocalizations. Three such judges were used all of whom were naive to the purpose of the ratings. They were presented with a sheet of instructions and asked to rate each voice on the following three dimensions: 1) pleasant-unpleasant, 2) warm-cold and

---

* However, only 58 voice samples were rated. Two were missed due to machine failure.

3) personal-professional*. A 10 point scale was used for each dimension (see Appendix N). Raters worked through the series of 58 voices with several rest periods.**

Each dimension was tested using an analysis of variance design with the usual five treatments as one factor and the three points in time as the other. No significant effects were found for the pleasant-unpleasant and warm-cold dimensions (See Tables 8 and 9, Appendix O). However, as Table 7 indicates for the personal-professional dimension, the main effect of treatments and treatments X time interaction were significant, The Tukey method for group comparisons was then applied yielding ten comparisons. The only comparison exceeding the critical ratio was that between the positive-negative and negative-positive groups. The mean for the positive-negative group was 19.33 while that for the negative-positive was 14.75 yielding a difference of 4.58 which exceeded the critical ratio of 4.07 ($\underline{P}$= .05). The interaction term was tested for

---

* The first two dimensions were chosen because of their relevance to reinforcement theory. Scheurer (1971) has presented evidence indicating that the pleasant-unpleasant and warm-cold dimensions are sufficiently independent to warrent considering them separately. The last dimension was included since Rosenthal (1966) has indicated that the most effective influence agents tend to be professional in manner.

** Fisher's z' statistic was used to average the pearson correlations among the three raters. On the pleasant-unpleasant dimensions z'=.28; for warm-cold dimension z'=.14 and on the personal-professional dimensions z'=.08. The respective r's were .27; .12 and .08. Using the Spearman-Brown formula to generalize to the three raters, the values become: .52 for the pleasant-unpleasant dimension; .29 for the warm-cold dimension; and .20 for the personal-professional dimension.

all possible pairwise comparisons. The only difference of significance was that between two cells: the positive-negative experimental group for the "middle" time slot and the average-average group for the "end" time period. This analysis is limited, however, by the low inter-rater reliability previously reported.

Self-Esteem Inventory and Coding Test

By ANCOVA with the Self-Esteem Inventory as the co-variate, no significant main or interaction effects of treatments and sections were obtained for the post scores on the Self-Esteem Inventory (See Table 10, Appendix O). Non-significant results were also obtained for the analysis of individual sub-scales of the Inventory (See Table 11, Appendix O).

Table 12, Appendix O, presents the ANCOVA for the coding test; the covariate was the coding pre-test. There was no significant main effect of treatments but a significant main effect of sections ($P = .04$). Examination of the means indicates no theoretically relevant differences. The Tukey method found that no pariwise comparisons were significant.

Related Findings

Both teachers and students had been asked to indicate in a Yes/No fashion whether or not the expectation feedback was acceptable to them. For the total sample, the teachers accepted 72% of the cases and the students 71%. ANCOVA of the Iowa scores, with the Iowa pre-test as the covariate, in a treatments X acceptance (Yes/No) design, for the teachers

and students separately, indicated no significant main or interaction effects.

Measures were taken during the post-testing session of the number of minutes each teacher took to administer the battery of tests. None of the differences between treatments and teachers were significant.

IX. Discussion

This research was planned to extend and improve methodological procedures of past studies in this area. One criticism has been that in several studies the measure of the dependent variable was of questionable reliability. Hence, changes in pre to post test scores may have been due partly to the unreliability of the instrument rather than to the effects of the experimental induction. For example, Snow (1969) in his review of Rosenthal's initial work (1968) cited evidence from the study itself that (see page 16) questions the reliability of the dependent variable measure (Test of General Ability - TOGA). (TOGA has been used as the dependent measure in seven of the thirteen studies reviewed earlier). In terms of validity, both the test publisher, Flanagan (1960), and Rosenthal (1968) view TOGA as a measure of intelligence. However, in TOGA's test manuel, correlations with 19 other measures of intelligence produced a median coefficient of .62; the correlations ranged from .41 to..80. In addition, since the test was not standardized on younger children, it may not be appropriate for a younger age group.

And in Rosenthal's 1968 study, it was the younger age groups which showed the greatest expectancy effect.

In contrast, the Iowa Test of Basic Skills was chosen for the current study because of its history of reliability as a measuring instrument (see page 39). Also this test has been standardized with younger age groups and hence is more appropriate for them.

A second criticism mentioned by Snow (1969) and Thorndike (1968) is the failure of some studies to cope with the problems raised by repeated measures of the same subjects. For example, TOGA was administered up to four times in some expectancy studies; and the results reported without consideration of the problem of repeated measures. In the current study, equivalent forms of the Iowa Test of Basic Skills, are provided by the test publishers, and were used.

A third criticism relates to the standardization of the pre-testing procedure. For example, Snow (1969) in reviewing Rosenthal's (1968) work noted that TOGA "...was administered to separate classes by the teachers themselves; this adds considerable uncertainty about standardization of procedure." In the current research, E's accomplice administered all pre-tests in all four sections thereby insuring more uniform administration and procedure.

A fourth area arises from lack of evidence that the experimental induction was successfully implemented. For example, Rosenthal's (1968) induction procedure entailed giving each teacher a list of "bloomers", following a

faculty meeting. No evidence is supplied that the teachers read the memorandum. In addition, no discussion is presented of the teachers' reactions to the "bloomers" and whether they found the list acceptable or incredulous. Following the experiment, Rosenthal asked teachers to report on the "bloomers" in his/her class; they were unable to do so. When presented with a list of four pupils, two of whom were "bloomers" and two control subjects who were matched on sex and age, the teachers were still unable to correctly select the "bloomers." These outcomes throw serious doubt on the effectiveness of Rosenthal's induction procedure. Other expectancy studies suffer from similar difficulties (Meichembaum, 1968; Anderson and Rosenthal, 1968; Clairborn, 1968; and Conn, Edwards, Rosenthal and Crowne, 1968).

In the current research it was considered essential that the induction process be clear and unequivocal and that the subjects reactions to the expectancy feedback be measured. Appendixes K and L indicate the "report form" handed to teachers and students which were read by all subjects in the presence of the E or his accomplice. In addition, the subjects indicated their acceptance of the experimental induction by circling a Yes or No on the form, in answer to a question about their reaction to the expectancy feedback.

In past research, teachers were given different expectations for different groups of students within the classes. With class sizes of 20-30 students, teachers probably found it difficult to remember how each student had performed on

the "blooming" test. In this study, post-testing by the teacher occurred in small groups of 4 to 6 students, all of whom had the same expectancy feedback. This homogenous grouping by expectancy feedback provided clarity to the teachers; and the small group atmosphere enhanced the opportunity for a close exchange between teacher and student.

Also, the interval between the administration of the induction procedure and the post-testing varied over previous expectancy studies. During lengthy time intervals, teachers may have had second thoughts about the validity of the expectancy feedback for certain pupils or have changed their expectations on the basis of further experience with the pupils. To minimize such occurrences in the current research the interval was held to 5-10 minutes.

A fifth concern relates to the failure of past research to consider both sides of the E-S dyad. Expectations were given only to the teachers, not to the students. To my knowledge this study is the first to create expectancies in both student and teachers. Further, Meichenbaum et al (1968) have suggested that both teachers' and students' pre-experimental expectations interact with experimentally induced ones and hence should be taken into account in designing research in this area. In the current study both teachers' and students' pre experimental expectations (see Appendixes E and F) were obtained, permitting the equalization of pre-experimental expectations across treatment conditions.

Another improvement in research design was to create negative expectancy feedback in subjects. With the exception of two studies reviewed earlier (Beez, 1968; Brager, 1969), all others have only used positive expectancy feedback. Following the line of reasoning outlined by Argyris (1969), in which he argues for a greater degree of "ecological validity", negative and expectancy feedback was included in order that the usual classroom setting would be simulated.

Finally, previous research has not attempted to measure the student's feelings about himself in relation to expectancy feedback. For example, various types of feedback may have implications for the student's self-esteem, separate from their effects on the main dependent variable. Hence, Coopersmith's (1967) Self-Esteem Inventory was included in the battery of pre and post tests in this research.

The preceding comments concerning the design of past research and the current study have been made to provide perspective for evaluating the significance of the current data. The results should be viewed in relation to both the experimenter bias effect and the operation of the reinforcement explanation of the effect.

The reinforcement approach to conceptualizing the mediating factors in the experimenter bias effect suggested three areas of analysis of the data from this experiment: examination of the predictions of direction of change in the five treatment conditions; comparison of scores on the first and second halves of the Iowa Test; examination of the

content of the teachers' verbalizations. In regard to all expectations (see page 33), the data are non-supportative.

Based on reinforcement theory, it was predicted that the positive-positive and negative-negative groups would show maximum opposite shifts on the dependent variable, with the former increasing and the latter decreasing. With both teacher and student expecting positive or negative student performance, the complex of covert cues in the dyad should reinforce each other's positive or negative expectations and produce maximum movement on the dependent variable. The findings indicated that these groups changed in a direction opposite to the predictions. For example, on the net proportion who change test, 26% of the negative-negative group improved on Iowa score from pre to post testing; in marked contrast, a net of 35% of the positive-positive group showed a decrement in Iowa score from pre to post testing. Likewise, with the sign test it was found that 64% of the scores for the negative-negative group increased; for the positive-positive group 68% went down. Also, the ANCOVA of post-test means (see Table 1) indicates no significant difference between these groups providing further evidence for lack of support for these predictions. In fact, the four treatment conditions containing various degrees of positive or negative feedback did not differ significantly from each other.

Examination of the results from the analysis of first and second halves of the Iowa test also indicates lack of

confirmation of the predictions. Reinforcement theory suggests that as the teacher-student interaction continues, they should shape each other's behavior through differential reinforcement. It is useful to review here, as an example of the operation of this phenomenon in the classroom, the studies of Canavan (1968) and Gumpert and Gumpert (1968). Given positive teacher expectation, a minor change in the student's behavior, say a nod or smile, may be interpreted as a sign of underatanding and the teacher thereby encouraged may intensity his efforts to teach the child. Subtle changes in the teacher's behavior such as alterations in body posture, Facial expression, vocal expression and praise may encourage the student and lead to increased interest and motivation and hence learning. This increased interest and learning then acts to fulfill the teacher's positive expectations and in this way the teacher's and student's behavior are shaped through mutual reinforcement. A similar process would be indicated for the negative expectations. Following this line of reasoning, responses later in the teacher-student interaction should be more in line with induced expectations than earlier ones. On the basis of the item difficulty data presented by the test publishers, both forms of the tests were divided into comparable halves. The results of this split half analysis presented in Tables 5 and 6 indicate no significant change in responding between the halves of the Iowa test for all of the five treatment conditions. A repeated measures ANCOVA (see Tables 6 and 7, Appendix O) confirmed the absence of

any significant main or interaction effect.

The tape analysis was designed to illuminate the content of delivery of some reinforcing stimuli. Previous research (Carlson and Hergenhahn, 1968; Moffatt, 1966) had indicated that the auditory channel is critical in the mediation of expectancy effects; the subjects had no access to the auditory cues and no significant expectancy effect was found. For the three dimensions measured in the current research, personal-professional, warm-cold and pleasant-unpleasant, the predictions were comparable to those indicated previously for the Iowa test. For the positive-positive group the prediction was that the teacher would be rated as speaking in a warmer and more pleasant manner in comparison to the negative-negative group. In addition, it was predicted that vocal utterances by the teacher should become warmer and more pleasant as the post testing session progresses. A reverse pattern of responding over the course of the post-testing session should prevail for the negative-negative group.

The predictions were not borne out. However low inter-judge reliability preclude placing much confidence on the tape analysis including the sole significant effect found on the personal-professional rating. It is possible that the inter-judge reliability problem is itself an indication of the absence of an experimenter bias effect.

In general, the results can be interpreted not only as a disconfirmation of the predictions set forth in this research but also as casting further doubt on the validity

of the experimenter bias phenomenon. Barber and Silver (1968) have raised serious questions about the validity of the phenomenon in laboratory reserach (see page 3), while Snow (1969) and Thorndike (1968) have raised similar questions concerning field studies (see page 16). Attempts to replicate significant findings have at times failed. For example, Barber (1969) failed to replicate the experimenter bias effect using 51 experimenters. Likewise, in field research, several attempts to replicate the initial Rosenthal (1968) study have failed to yield similar results (Anderson and Rosenthal, 1968; Conn, Edwards and Crowe, 1968.

The significance of the current disconfirmation should be judged in relation to the authors' pre-experimental expectation in favor of the effect and the utilization of comparison conditions that should have facilitated the finding of an effect, if the phenomenon were real. In the few previous studies where both positive and negative feedback were employed- a strong main effect was found; stronger than when only positive feedback was employed. Beez (1969) had induced both favorable and unfavorable expectations in teachers working with Head Start children. Rosenthal, (1969) in commenting on Beez's procedure, noted that including both positive and negative feedback gave a better estimate of expectancy effects as they exist in the real world of the classroom where negative expectations doubtless exist. In the current research, in addition to including negative expectation feedback, both sides of the teacher-student dyad in two treatment conditions

received similar feedback. Hence,the likelihood of obtaining significant differences between the four treatment conditions containing positive or negative feedback should have been enhanced by these procedures. Consequently, the lack of significant differences as evidenced in examination of Table 1, reflects negatively on the presence of the experimenter bias phenomenon.

Other aspects of this design also were seen as enhancing the likelihood of obtaining significant differences between these four treatment conditions. The average post testing session lasted 47 minutes. Earlier, evidence was presented that the induced expectancy is communicated from E to S in the very first moments of their contact. Rosenthal (1966) has summarized a body of data suggesting that expectancies are communicated during the brief period when the E greets, seats, and instructs his subject. The point being made here is that there was sufficient time during the post test for the communication of expectancies to occur and that on this basis absence of significant differences between these four treatment conditions casts further doubt on the reality of the experimenter bias phenomenon.

In terms of the reinforcement literature reviewed earlier (see page 6, 19 and 22), the current results confirm the findings of Rosenthal (1966) and Weick (1965) who employed the classic photo rating task and found no change in responding as the experiment progressed. Reinforcement theory suggests that the very first response of a given subject is not

affected by the experimenter's expectancy and that later responses are more affected than earlier ones. Rosenthal (1966) plotted the ratings in sequence and found that the magnitude of the experimenter expectancy effect "changed very little over time."

Brager (1969) has presented evidence that the formality or informality of the E-S interaction is an important factor in facilitating the experimenter bias effect. He varied the task and consequently the level of E-S interaction along a formality-informality dimension. His tasks varied from taking a mathematics test (low interaction-high formality) to the reading of a story to students who then commented on the story (high interaction-low formality). A significant experimenter bias effect was found only when the tasks allowed considerable E-S interaction. The task variable employed in the current research (i.e., reading test items) may have structured the E-S interaction in such a way as to discourage informal interaction and make the situation similar to brager's high formality condition. Hence, the lack of positive findings in the current research may be taken to confirm Brager's assumption that if the experimenter bias effect appears at all, it is less likely to appear with highly structured tasks.

Earlier (see page 9) it was indicated that experimenter bias effects may be a learned phenomenon, i.e., one that is learned in interaction with a series of research subjects. It was concluded then, based on a review of twelve studies,

that subjects seen later in an experiment showed a greater expectancy effect than those seen earlier. This essentially is what Meichenbaum et al (1966) accomplished in structuring a series of E-S interactions (i.e., class sessions); and his work supported a reinforcement position (see page 20). In the current research, teachers conducted ten post-testing sessions but they were counterbalanced as to expectancy feedback (see Appendix M; Tables 3 and 4) to control sequence effects and the effect of uncontrolled events. Hence, the current research did not lend itself to this type of analysis. The examination of the reinforcement position as it relates to a series of E-S interactions is an area recommended for further research.

There are a variety of speculations that can be cited as explanations for the lack of support for the predictions from this study. One is that the experimental induction failed. The results of the Yes/No analysis of subject responses to a question about their acceptance of the induction procedure does not support this view. As reported earlier (see page 54), 72% of the teachers and 71% of the students said they accepted the expectancy feedback. Also, ANCOVA in a treatment X acceptance (Yes/No) design indicated no significant differences between the treatments whether or not the expectancy was assertively accepted by teachers or students.

Another speculation is that the interval between the experimental induction and the post-test was too brief for

the full impact of the induction process to appear. It was indicated earlier that within a laboratory setting the appearance of the expectancy phenomenon is swift; however these findings were based on simple tasks (i.e., photo ratings). In the current research the area of expectancy feedback had far greater import for the subjects' academic performance. Perhaps in an area reflecting on students' academic ability, teachers need more time to incorporate the expectation and reduce whatever dissonance may be aroused. Future studies might employ measures designed to aid the teacher in accepting the expectancy feedback. For example, Meichenbaum et al (1968), discussed each of the "bloomers" with the teachers. In fact, previous teachers of the "bloomers" attended the session and offered support for the assertions of the teachers. They report that by the end of the session teachers were showing signs of incorporating the expectancy feedback despite their doubts.

In short, the combination of insufficient time for the teacher to incorporate the expectancy feedback and a task area reflecting a stable area of student functioning may, in combination, have militated against the predictions of the study.

Finally, we will take up the matter of the significant difference between the average-average condition and the four other treatment conditions containing positive and/or negative feedback. It will be recalled that a pure control group devoid of expectancy feedback was impossible given the need

to have all treatment conditions in all sections (see page 31). The closest approximation to a control group under these circumstances was to provide average or neutral feedback. The significant improvement in this group was unexpected and unpredicted. To better understand the change in this group, the mean post-test scores for subjects scoring above and below the pre-test mean on the Iowa test and the Self-Esteem Inventory were computed (see Table 17, Appendix O).* It is evident that the overall gain on the Iowa test by the average-average group was due to the increment of those initially below the mean. Since similar magnitudes or directions of change were not exhibited by students below the pre-test means in the other groups, regression alone can not explain the results. Also, this result cannot be due to a correlated change in self-esteem, as the Tables reveal. However, no sound explanation is available for this result, which of course may merely be a chance finding.

---

* Similar Tables are prepared for the four other treatment conditions (See Appendix O; Tables 13, 14, 15 and 16). There was no apparent pattern to the results of these computations.

TABLE 1

Mean, standard deviation, change scores and adjusted means for treatment factor.

Iowa Test of Basic Skills.

	Pre-Test		Post-Test		Change on Mean	Covariance Adjusted Mean
	Mean	Standard Deviation	Mean	Standard Deviation		
NEGATIVE-NEGATIVE (N=19)	49.87	9.5	49.28	11.0	- .59	49.31
POSITIVE-POSITIVE (N=20)	50.06	10.8	48.11	10.6	-1.95	48.02
POSITIVE-NEGATIVE (N=21)	50.27	9.6	48.95	10.1	-1.32	48.70
NEGATIVE-POSITIVE (N=19)	49.68	10.0	49.38	10.5	- .30	49.59
AVERAGE-AVERAGE (N=21)	49.85	11.6	53.75	8.7	+3.90	53.83

Table 2

Analysis of Co-variance For Five Treatment Groups X Four Sections

	d.f.	Mean Square	F
TREATMENTS	4	102.4281	2.46[a]
SECTIONS	3	154.8477	3.72[b]
TREATMENT X SECTIONS	12	29.9775	0.72
REGRESSION	1	5333.2187	128.35
ADJUSTED ERROR	79	41.5501	
ERROR	80	107.6961	

a. P= .05

b. P- .01

TABLE 3

Net Proportion Who Show An Increase In Iowa Scores

GROUPS	Net Proportion
NEGATIVE-NEGATIVE	+26%
POSITIVE-POSITIVE	-35%
POSITIVE-NEGATIVE	-5%
NEGATIVE-POSITIVE	-6%
AVERAGE-AVERAGE	+42%

TABLE 4

Intercorrelation Matrix For Entire Sample

(N=100)

	Teachers Pre	Students Pre	SEO Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.41	-.20	.24	.79	40.65	6.73
Teachers Pre-Expectations		-.10	.06	.35	5.09	3.23
Students Pre-Expectations			-.47	-.29	1.66	.60
SEI Pre-Test				.22	65.64	15.26
Iowa Post-Test					41.37	8.11

Table 5

Mean Score on Iowa Post-Test Items Re-grouped into Halves by Treatment Group - Fifth Grade

TREATMENT GROUP	MEAN SCORE FIRST 31 ITEMS	MEAN SCORE SECOND 31 ITEMS	CHANGE
NEGATIVE-NEGATIVE	25.00	24.00	-1.00
POSITIVE-POSITIVE	22.16	22.16	-
AVERAGE-AVERAGE	24.66	25.66	+1.00
POSITIVE-NEGATIVE	24.40	24.60	+ .20
NEGATIVE-POSITIVE	23.33	23.33	-

Table 6

Mean Score on Iowa Post-Test Items Re-grouped into Halves by Treatment Group - Fourth Grades

TREATMENT GROUP	MEAN SCORE FIRST 26 ITEMS	MEAN SCORE SECOND 26 ITEMS	CHANGE
NEGATIVE-NEGATIVE	20.07	18.14	-1.92
POSITIVE-POSITIVE	19.07	19.14	+ .07
AVERAGE-AVERAGE	20.93	20.46	- .46
POSITIVE-NEGATIVE	19.06	19.33	+ .25
NEGATIVE-POSITIVE	19.78	18.35	-1.32

Table 7

Analysis of Variance of Personal-Professional Dimension.
3 (Beginning-Middle-End) X 5 (Treatment Groups)

	d.f.	Mean Square	F
Beginning-Middle-End	2	10.76	.90
Treatment Groups	4	36.63	3.06[a]
Beginning-Middle-End X Treatment Groups	8	39.57	3.31[b]
ANOVA Error	44	11.95	

a. P= .02

b. P= .005

X. LIST OF APPENDIXES

APPENDIX A: Tables summarizing past studies.

APPENDIX B: Excerpts of pre and post tests of Iowa Test of Basic Skills - Grades 4 and 5.
B1 = Iowa pre-test for Grade 4.
B2 = Iowa post-test for Grade 4.
B3 = Iowa pre-test for Grade 5.
B4 = Iowa post-test for Grade 5.

APPENDIX C: Coding sub-test of WISC.

APPENDIX D: Self-Esteem Inventory (excerpts).

APPENDIX E: Teachers pre-experimental expectation measure.

APPENDIX F: Students pre-experimental expectation measure.

APPENDIX G: Memorandum to teachers.

APPENDIX H: Explanation sheet to teachers.

APPENDIX I: Letter of explanation to parents.

APPENDIX J: Explanatory sheet to students.

APPENDIX K: "Report forms" to students.

APPENDIX L: "Report form" to teachers.

APPENDIX M: Detailed account of procedure.

APPENDIX N: Material used in voice analysis.
N1 = Instructions for voice raters.
N2 = Attributes rating form.

APPENDIX O: Appendix tables.

APPENDIX A

Table 1

Summary of studies dealing with mediating variables in laboratory research

		Study	Sign. level
Variable:	Operant Conditioning (4 studies)	Rosenthal (1966)	N.S.
		Masling (1965)	N.S.
		Weick (1965)	N.S.
		Adair & Epstein (1967)	N.S.
Variable:	Communication Channel (8 studies)	Adair & Epstein (1967) (Used auditory channel)	Z = +1.64
		Troffer & Tart (1964) (Used auditory channel)	Z = +2.81
		Rosenthal & Fode (1963) (Used auditory channel)	47% of variance
		Duncan & Rosenthal (1968) (Used auditory channel)	Sign.
		Carlson & Hergenhahn (1968) (Used auditory channel)	Sign.
		Moffatt (1966) (Used auditory channel)	Sign.
		Zolbe (1968) (Used both auditory and visual channels)	Visual more effective Z = +1.44
		Kennedy, Edwards & Winstead (1968) (Used visual channel)	Sign.

Variable:	Interpersonal learning (13 studies)	3 Rosenthal studies (1966)*	Sign.
		3 Rosenthal studies (1966)*	N.S.
		2 Uho, Frager & Rosenthal studies (1968)*	Sign.
		2 Uho, Frager & Rosenthal studies (1968)*	N.S.
		Conners & Horst (1966)	Sign.
		Canavan (1968)	Sign.

*Summarized in Rosenthal (1969)

Variable:	Organismic characteristics of E - Sex (7 studies)	4 studies summarized by Rosenthal (1969)	Sign. Male more effective
		Johnson (1967) E & S same sex	Sign.
		Adair (1968) E & S same sex	N.S. Z = -2.33
		Silverman (1968)	N.S. Z = -1.61
Variable:	Psycho-social characteristics of E - Dominance - Professional cluster. (10 studies)	5 Rosenthal (1966) studies using direct observation	Cluster analysis sign. for Dom-Prof. cluster
		5 Rosenthal (1966) studies, using film observation	Cluster analysis sign. for Dom-Prof. cluster

APPENDIX A
Table 2

Teacher Expectancy Effects in 13 Experiments

Study	Criterion	Results p	Results z
Anderson & Rosenthal (1968)	Total I.Q.	--	0.00[a]
Beez (1968)	Symbol learning	.000002	+4.67
Biegen (1968)	Achievement	.002	+2.89[b]
Burnham (1968), Burnham & Hartsough (1968)	Swimming score	.005	+2.61[a]
Clairborn (1968)	Total I.Q.	(-).08	-1.45[a]
Conn, Edwards, Rosenthal & Crowne	Total I.Q.	---	0.00[a]
Flowers (1966)	Achievement + I.Q.	.06	+1.60
Heiserman (1967)	Aspiration	---	0.00
Meichenbaum, Bowers & Ross (1968)	Classroom behavior	.02	+2.02[b]
Pitt (1956)	Achievement	---	0.00
Rosenthal & Anderson (1969)	Total I.Q.	(-).06	-1.56[b]

Rosenthal & Evans (1968)	Total I.Q.	---	0.00[a,b]
Rosenthal & Jacobson (1968)	Total I.Q.	.02	+2.11[a]
	Overall	.0002	+3.57

a = Indicates the interaction of teacher expectancy with some other variable.

b = Preliminary data.

APPENDIX A
Table 3

Summary of studies dealing with mediating variables in field research

Study	Mediating variable	P value
Rosenthal & Jacobson (1968)	Differential time	N.S.
Meichenbaum (1968)	Differential time	N.S.
	Interaction quality	Sign.
Conn, Edwards, Rosenthal, & Crowne (1968)	Auditory cues	Sign.
Beez (1968)	Amount of material taught to expectancy group.	Sign.

APPENDIX A
Table 4

Summary of Recent Studies in Expectancy Research

Study	Task and Criterion	Type of variables investigated and theoretical orientation	Results
Kessell & Barber (1968)	Verbal conditioning in laboratory	1) Physical and personality characteristics of E. 2) S's attraction to E on the basis of S's interaction with him. 3) S's attraction to E manipulated on the basis of matching on psychological test scores. 4) S's attraction to E manipulated by having collaborator give S information about E. Theoretical framework was interpersonal attraction and influence.	Conflicting. Authors conclude that E charact. play a role but in interact. with other variables in complex fashion.
Kennedy (1969)	Taffel-type verbal conditioning task in laboratory setting	1) Three levels of E expectancy: positive, neutral and negative. 2) Mode of stimulus presentation. Theoretical framework was reinforcement paradigm.	N.S.

Sheehan (1969)	Verbal conditioning in laboratory by constructing sentences.	Manipulated E's expectancy of role of awareness in conditioning.	Sign. only for those S's and E's who were unaware.
Brager (1969)	Students rated a story and picture and took math test.	1) Level of E-S interaction; from structured to unstructured. 2) Level of E expectancy: positive, negative & neutral.	Sign. for tasks allowing maximum E-S interaction.
Kester (1969)	1) I.Q. scores. 2) Achievement test scores. 3) Measure of teacher and student interaction. 4) Attitude measure. Setting was classroom.	1) Pupils' attitudes. 2) Quality of E-S interaction. 3) Time spent with pupil. 4) E-S feedback loop. Theoretical framework was dyadic interaction model.	Of ten hyphot: 3 N.S. 3 partly supported 4 Sign.
Haskett (1969)	1) Scores on Metro. achiev. tests. 2) Syracuse Scale of Social relations. 3) Setting was classroom. Videotapes taken and measure of time devoted by teacher to student.	1) Academic achievement. 2) Area of social development. 3) Rate of achievement. 4) Sex of subject. Theoretical framework was E-S interaction model.	Sign. for social develop. and academic areas.

Barber et al (1969)	Ratings of photos in classroom setting.	Induced positive and negative expectancies in E.	N.S.
McFall (1970)	Standard photo-rating task in laboratory setting.	1) Need for achievement. 2) Field dependence-independence. 3) Theoretical framework was communication model of expectancy communication.	Sign. for high N-Ach & field dependent S's.
Jacob (1970)	Varied tasks from high to low ambigty.	1) Manipulated E's "demand characteristics" and 2) "investment" by E in results. 3) Manipulated ambiguity of task to predict type of mediating mechanisms.	N.S. S's were suspicious and there was low acceptance of exper. induction.

Excerpts from Iowa Test of Basic Skills: pre-test form 4 for grade four. Reprinted by permission of Houghton-Mifflin Company, Boston, Massachusetts.

APPENDIX B1
HUNTER TEST OF STUDY HABITS

Instructions: In each of the exercises below, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy BLACK CAPITAL type above them. Circle the correct answer below. The sample exercise in the box below has already been marked correctly.

Sample exercise

0. A TALL building
   1. High
   2. Wide
   3. Low
   4. New

1. A FAIRY story
   1) ghost
   2) make-believe
   3) true
   4) real life

. . .

Instructions: In each of these exercises, you are to choose the word that would appear first if the four words were arranged in alphabetical (a-b-c) order. Circle the correct answer.

24. 1) queer
    2) ram
    3) propose
    4) request

Below is the TABLE OF CONTENTS of a book called Know the Flag. Use this table to answer the exercises below it. Circle the correct answer.

TABLE OF CONTENTS

36. In which chapter might there be a picture of the flag of Texas, the Lone Star State?
    1) 3 3) 8
    2) 4 4) 9

40. Which chapter might tell about the rattlesnake flag that was carried by some American soldiers in the war for Independence?
    1) 1 3) 4
    2) 3 4) 8

Excerpts from Iowa Test of Basic Skills: p_{st}-test form 1 for grade four. Reprinted by permission of Houghton-Mifflin Company, Boston, Massachusetts.

APPENDIX B2
HUNTER TEST OF STUDY HABITS

Instructions: In each of the exercises below, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy BLACK CAPITAL type above them. Circle the correct answer. The sample exercise in the box below has already been marked correctly.

Sample exercise

0. A TALL building
   1. (High)
   2. Wide
   3. Low
   4. New

1. To TWIST the handle
   1) turn
   2) lift
   3) pull
   4) push

. . .

Instructions: In each of these exercises, you are to choose the word that would appear first if the four words were arranged in alphabetical (a-b-c) order. Circle the correct answer.

24. 1) habitual
    2) girl
    3) effort
    4) iceman

Below is the TABLE OF CONTENTS of a book called Food For All. Use this table to answer the exercises below it. Circle the correct answer.

TABLE OF CONTENTS

36. Which chapter might tell how much milk we should drink each day?
    (1) 5 (2) 6 (3) 8 (4) 9

40. Which chapter might tell us how food gets from the farmer to people in cities?
    (1) 3 (2) 4 (3) 6 (4) 7

Excerpts from Iowa Test of Basic Skills: pre-test form 4 for Grade five. Reprinted by permission of Houghton-Mifflin Company, Boston, Massachusetts.

APPENDIX B3
HUNTER TEST OF STUDY HABITS

Instructions: In each of the exercises below, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy BLACK CAPITAL type above them. Circle the correct answer below. The sample exercise in the box below has already been marked correctly.

Sample exercise

0. A TALL building
   1) (High)
   2) Wide
   3) Low
   4) New

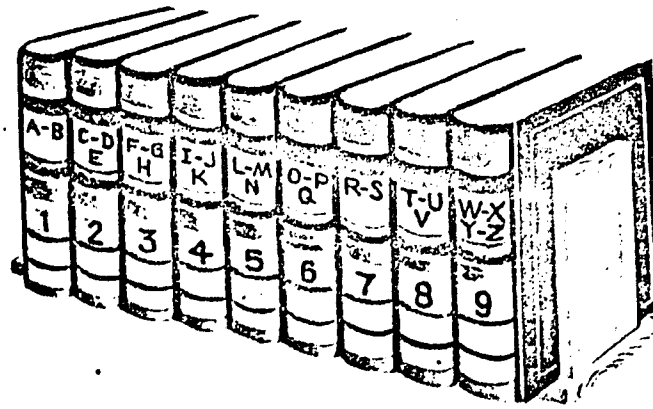
1. A iron BENCH
   1) fence
   2) seat
   3) bucket
   4) grill

. . .

Instructions: In each of these exercised, you are to choose the word that would appear first if the four words were arranged in alphabetical (a-b-c) order. Circle the correct answer.

6. 1) queer
   2) ram
   3) propose
   4) request

Below is a picture of a set of volumes of an encyclopedia. Each volume contains information about topics which begin with the letters shown on the back. Use the picture to answer the exercises below.



38. Which volume would contain a discussion of how X-rays are used in medicine and industry?
   1) Vol. 4 3) Vol. 7
   2) Vol. 5 4) Vol. 9

39. In which volume would you look first to see whether to see whether such tropical birds as the toucan and parro are found in North America?
   1) Vol. 1 3) Vol. 6
   2) Vol. 5 4) Vol. 8

Excerpts from Iowa Test of Basic Skills: post-test form 1 for Grade five. Reprinted by permission of Houghton-Mifflin Company, Boston, Massachusetts.

APPENDIX B4
HUNTER TEST OF STUDY HABITS

Instructions: In each of the exercised below, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy BLACK CAPITAL type above them. Circle the correct answer below. The sample exercise in the box below has already been marked correctly.

Sample exercise

0. A TALL building
   1) (High)
   2) Wide
   3) Low
   4) New

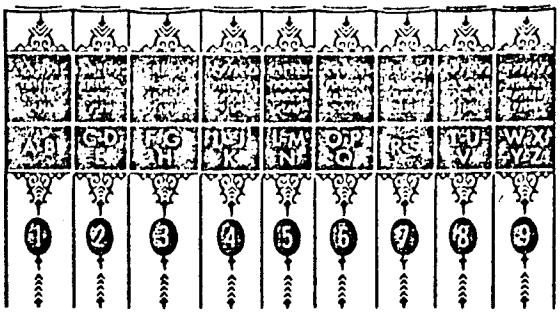
1. To MEND the fence
   1) damage
   2) jump
   3) build
   4) repair

.
.
.

Instructions: In each of these exercises, you are to choose the word that would appear first if the four words were arranged in alphabetical (a-b-c) order. Circle the correct answer.

6. 1) habitual
   2) girl
   3) effort
   4) iceman

Below is a picture of a set of volumes of an encyclopedia. Each volume contains information about topics which begin with the letters shown on the back. Use the picture to answer the exercises below.



38. If you wanted to read about the city of Rome, the capital of Italy, which volume would you choose?
   1) Vol. 2 3) Vol. 7
   2) Vol. 4 4) Vol. 8

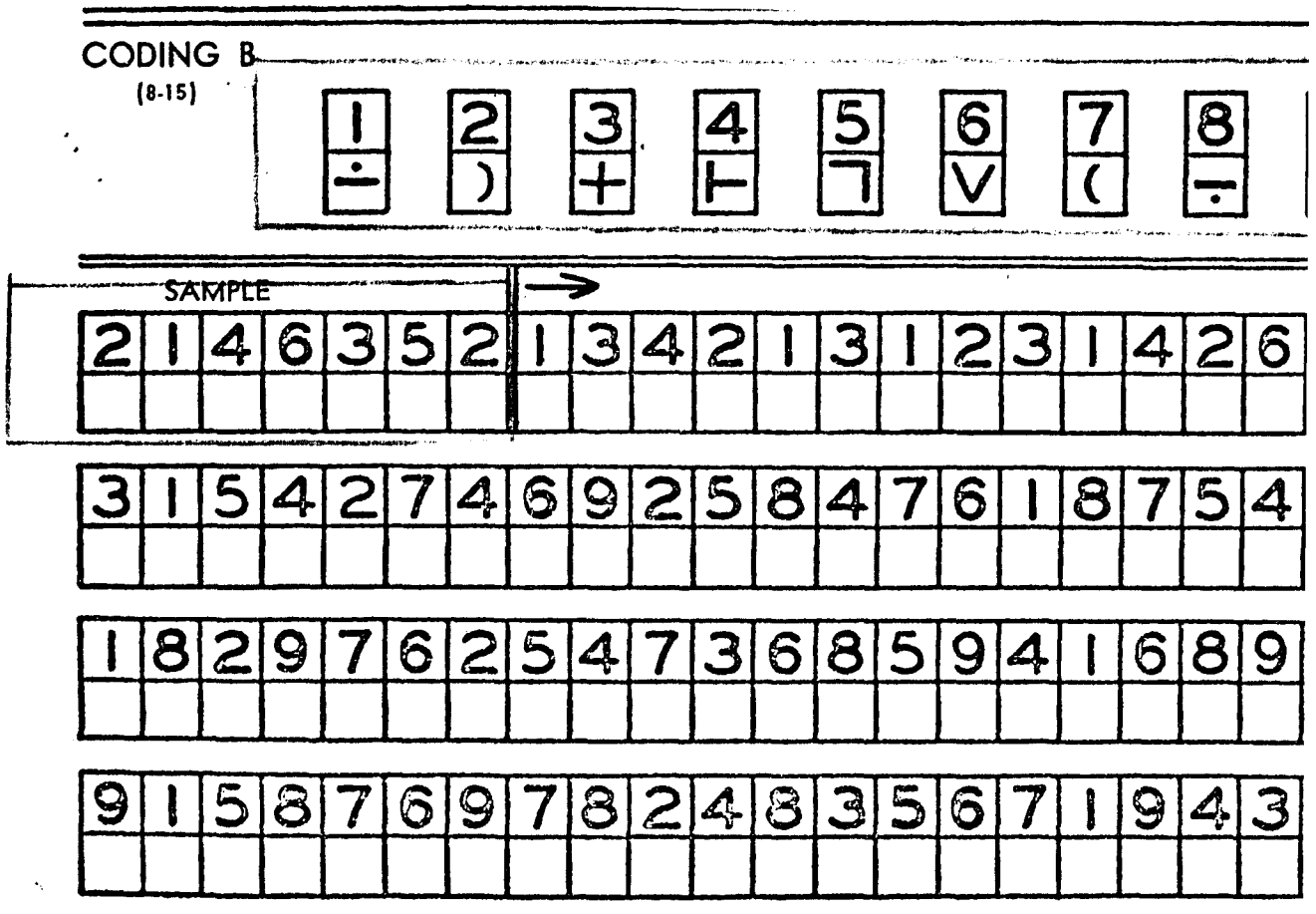
39. Arizona and Utah are among the leading states in copper production. Which volume would tell about some of the uses of this mineral?
   1) Vol. 1 3) Vol. 5
   2) Vol. 2 4) Vol. 6

APPENDIX C*
CODING TEST

Instructions: This is a kind of number and sign puzzle. It has to do with matching certain numbers with certain signs. Look at the set of boxes below that are outlined in red. This is the part of this test which shows you what to do. There is a number on the top part and a mark on the bottom part. Now look down to the next line and find the part that is outlined in green. This is a practice set. You are to find the mark that goes with the number and fill it in the correct place. Check back to the top set of boxes outlined in red to find the correct mark.

Let's do the green sample of boxes together to be sure we all understand these directions. The first number in the green sample is 2. Looking to the red boxes we see that the mark under 2 is a . Now you put a under number 2. All right, you may finish the green sample by yourselves.

Now, let's all find the black arrow. It is there right after your set of green boxes. You will begin to work where the black arrow is pointed. Fill in as many squares as you can without skipping any. All right --- BEGIN.



* Reprinted by permission of Psychological Corporation.

Excerpts from Self-Esteem Inventory used for all grades*

APPENDIX D

Self - Esteem Inventory

Please mark each statement in the following way:
If the statement describes how you usually feel, put a check (✓) in the column - "LIKE ME".
If the statement does not describe how you usually feel, put a (✓) in the column-"UNLIKE ME".

There are no right or wrong answers.

		LIKE ME		UNLIKE ME	
1.	I spend a lot of time daydreaming.				
2.	I'm pretty sure of myself.				
3.	I often wish I were someone else.				
4.	I'm easy to like.				
5.	My parents and I have a lot of fun together.				
6.	I never worry about anything.				
7.	I find it very hard to talk in front of the class.				
8.	I wish I were younger.				
9.	There are lots of things about myself I'd change if I could.				
10.	I can make up my mind without too much trouble.				
11.	I'm a lot of fun to be with.				
12.	I get upset easily at home.				
13.	I always do the right thing.				
14.	I'm proud of my school work.				

* Reprinted by permission of Stanley Coopersmith, Univ. of California at Davis.

APPENDIX E

HUNTER TEST OF STUDY HABITS

RATING FORM

Dear _____,

Here is a ten point scale on which you are asked to rate each child's study habits (i.e., his use of maps, dictionary etc.). Would you please indicate by placing an X over the correct numeral, where in your opinion, the child falls on the scale.

We want to emphasize that this should be a clear indication of your impression. The child's name is: _____

_____.

Scale

1	2	3	4	5	6	7	8	9	10
Poor				Average					Exceptional

_____: Check here if you have no opinion at this time.

APPENDIX F

HUNTER TEST OF STUDY HABITS

This is a test of study skills. It will help us to find out how good you are at such things as looking up words in dictionaries, alphabetizing, using an index and finding information. It has already been given to other 4th. and 5th. graders in other schools and it told the children how well they could use these materials. We are sure it will be able to tell you the same thing.

In a week or two, after we have checked everyone's papers, we will tell each of you how well you did. At that time we would like you to take the test over again.

Now, before the test starts we want to know how well you think you can alphabetize and use dictionaries and information books. Here is a set of 3 sentences. Please circle the one you think best tells how well you can use these materials:

DO YOU THINK YOU ARE:

(1) GOOD AT IT.

(2) FAIR AT IT.

(3) POOR AT IT.

Now turn the page and read along with your teacher.

APPENDIX G

GLEN HEAD SCHOOL
Sept. 10, 1970

TO: 4th Grade Teachers and Mrs. Lawrence
FROM: L. L. Marrone

Would you please plan on a brief meeting in my office on Thursday, Sept. 17th at 3:10.

Our district is participating in a research project on developing a test on Study Habits and other related areas. Our Board of Education approved the use of Glen Head School and Mr. Joseph Kern is the project director. He will be here next Thursday to explain your role in this study. We will make every effort to keep the process as simple as possible.

Louis L. Marrone,
Principal

LLM:rm
CC: Mr. Kern

APPENDIX H

HUNTER COLLEGE
OF THE CITY UNIVERSITY OF NEW YORK
695 Park Avenue
New York 21, N. Y.

HUNTER TEST OF STUDY HABITS

There has been much discussion among leaders in the field of education concerning the basic goals of the modern educational system. These discussions are particularly pertinent when one considers our rapidly changing and expanding society. Many of these Academicians feel that the only meaningful goal of today's school systems can be the development of the child's cognitive structure. This implies more and more independence on the part of the student in all fields of study which means the student must become increasingly more adept in the use of resource materials. In other words, the ability to learn how to study is crutial in adepting to the present day environment. With the rapidly increasing bodies of information available in all areas of learning it becomes increasingly more important for school districts to guide children at an early age in the development of adequate study habits.

Hunter College is very much interested in school children who show unusual proficiency in the use of resource materials, and has conducted classroom studies in this area for several years. A test has been devised to go along with these studies (THE HUNTER TEST OF STUDY HABITS) and its administration in

HUNTER COLLEGE
OF THE CITY UNIVERSITY OF NEW YORK
695 Park Avenue
New York 21, N. Y.

various schools throughout the country has revealed some interesting data. It would appear that children who demonstrate the ability to use these various materials effectively often do so in a sudden, identifyable spurt within the classroom setting. These spurts are demonstrated by hildren at all academic levels, and frequently by those who might seem least likely to possess this ability. In other words, just as there are "late bloomers" in other areas in the school situation, there would seem to be "late bloomers" in the ability to handle resource materials.

The Hunter Test of Study Habits, in view of the consistent results obtained in schools in the Mid-West and New England, would seem to be an accurate predictor of the child's ability in this area. We wish to further validate this test as a good predictor of this skill by administering it in your classroom. The classroom teacher will be asked to administer the test to her class as a whole. When the results are obtained, she will then be asked to re-administer the test on a small group basis, in order that the University can obtain a reinforcement of the original results. At that time, the teacher will also be advised of the results of the first testing.

The teacher will also be asked to administer two brief accompanying tests which are routinely given in conjunction

HUNTER COLLEGE
OF THE CITY UNIVERSITY OF NEW YORK
695 Park Avenue
New York 21, N. Y.

with the major one, to round out information on each child.

We hope this will be a meaningful experience for the teachers, in order to somewhat compensate for their generous assistance in time and effort.

APPENDIX I

NORTH SHORE SCHOOLS

Reply to: GLEN HEAD SCHOOL, 7 SCHOOL ST., GLEN HEAD, N.Y. 11545

LOUIS L. MARRONE
Principal
(516) 671-5500

Sept. 16, 1970

Dear Parents,

Glen Head School is participating with Hunter College of The City University of New York, in a brief educational research project concerned with understanding more about the learning process. The research will be conducted in our 4th and 5th grades and will take no more than about two hours of classroom time. It will be conducted by a graduate student, Joseph Kern, who is a candidate for a Doctor's degree at Hunter College.

The project has been approved by the School Board and School Administration. We are writing you to advise you why Mr. Kern will be in these classrooms.

Louis L. Marrone,
Principal

APPENDIX J

HUNTER COLLEGE
CITY UNIVERSITY OF NEW YORK

Boys and girls, today we are going to ask you to help us with some studying that we grown-ups are doing in school. You know, grown ups often keep going to school to learn things, just the way you are doing now. We are trying to learn more about boys and girls so that we can help them to do their best in their classrooms. One of the ways we learn about them is to ask them different kinds of questions about themselves and their work. We would like to ask you these questions this morning, and we are going to give you three different kinds of jobs to do. One is some questions about yourself - how you feel about certain things. Another is a kind of number and sign puzzle, and the third one will be some questions about how you use school materials. We think that your answers will help us to learn more about schools, and we also think they will help you in some ways, too.

Thank you very much for helping us do this.

During all three sets of questions, your teacher will read everything to you out loud. This is to help you. Be sure to follow along with the teacher and do not rush ahead of her. She will give you time to answer each question before she goes on to the next.

Now we will start with the first set of questions. Please turn the page and read along with your teacher.

Appendix K1 - Positive Feedback Group

HUNTER TEST OF STUDY HABITS

Report Form To Students

Dear_____,

Do you remember the test your teacher gave you when school first started? There were lots of questions for you and all the other children in your class to answer. Remember, your teacher told you that the test would tell how well you can use things like a dictionary, maps and information books?

Well, we've looked at your test and:

> You did very well. You certainly use these materials well. You are one of the best in the class.

- -

Do you believe that's true about yourself? Yes.....1
No.....2

Appendix K2 - Negative Feedback Group

HUNTER TEST OF STUDY HABITS

Report Form To Students

Dear _____,

Do you remember the test your teacher gave you when school first opened? There were lots of questions for you and all the other children in your class to answer. Remember, your teacher told you that the test would tell how well you can use things like a dictionary, maps and information books?

Well, we've looked at your test and:

You didn't do very well, I'm afraid. Most of the other children did better than you did.

- -

Do you believe that's true about yourself? Yes.....1
No.....2

Appendix K3 - Average Feedback Group

HUNTER TEST OF STUDY HABITS

Report Form To Students:

Dear_____,

Do you remember the test your teacher gave you when school first started? There were lots of questions for you and all the other children in your class to answer. Remember, your teacher told you that the test would tell how well you can use things like a dictionary, maps and information books?

Well, we've looked at your test and:

> You came out about in the middle of the class. Half of the class did better than you did, and half did not do as well as you did.

- -

Do you believe that's true about yourself?

Yes.....1
No.....2

Appendix L1 - Positive Feedback Group

HUNTER TEST OF STUDY HABITS

Dear _____,

The results of the Hunter Test of Study Habits, with regard to _____ are as follows:

> He demonstrates an exceptional ability to use resource materials. In fact, he is predicted to be one of your outstanding students in the use of resource materials in any projects you may be doing this year.

- -

Could you tell me: Do you accept this statement as reasonable?

Yes.....1
No.....2

Appendix L2 - Average Feedback Group

HUNTER TEST OF STUDY HABITS

Dear _____,

The results of the Hunter Test of Study Habits, with regard to _____ are as follows:

> His ability to use resource materials is about average. He fell approximately on mid-point on the test scale, with an equal number of students doing better and poorer than he.

- -

Could you tell me: Do you accept this statement as reasonable?

Yes.....1
No.....2

Appendix L3 - Negative Feedback Group.

HUNTER TEST OF STUDY HABITS

Dear_____,

The results of the Hunter Test of Study Habits, with regard to _____are as follows:

> He fell within the lowest score range on the test scale. You probably will not see much ability to use any kind of resource material from this child during the course of this school year

- -

Could you tell me: Do you accept this statement as reasonable?

Yes.....1
No.....2

APPENDIX M

Steps in Execution of Expectancy Study

Step 1: School Board Approves Research

The Superintendent of Schools and the principal whose school was the setting for the research reviewed the thesis proposal during the spring of 1970. They were the only two officials in the school district who were aware of the true intent of the experiment. Upon their recommendations, the school board in May 1970 passed a resolution approving of the research. At that time the school board was given a brief explanation and told that the purpose of the research was, in general terms, to "learn more about the learning process."

Steps 2 to 5: Meetings with School Principal and Teachers

On September 10, 1970, the E met with the school principal to work out details of the research. There were three sections of Grade 4, all of which were included in the design, as well as one section of Grade 5. The fifth grade class was selected from the others available because it was felt that the teacher represented the typical school teacher in the district and had displayed a cooperative manner in other areas.

It was decided that the pre-test would be administered by an accomplice of the E. The principal advised that, in general, the school did not administer achievement or group intelligence tests. It was felt that pre-test administration by the accomplice would provide a more standardized procedure. This will be discussed in more detail later.

It was also decided that the pre-testing would take place in the children's regular classroom. All post-testing, however, was to be completed in a small room in the school so that the

students would be exposed to a fixed stimulus environment.

On September 14, 1970, the school principal sent a memorandum to the teachers advising them of the research and asking them to attend a meeting with him and the E to work out details and explain their role in it (see Appendix G). On the same date at a school-wide faculty meeting, the principal announced that research would be conducted in these grades.

On September 17, 1970, a meeting was held with the four participating teachers in the principal's office. The purpose of this meeting was to orient the teachers to the study. The principal opened the meeting by telling the teachers that the research had the full support of the school administration and school board. He said that the school district was committed to educational research and was pleased to have this opportunity to be of help.

The principal then introduced the E. There were several points I made in my remarks: (1) the main point of the experiment was to gather further validity and normative data on the Hunter Test of Study Habits (actually this was the Iowa Test of Basic Skills). I pointed out that the test had been used successfully in New England and the Midwest to predict potential bloomers in the area of study habits. I said that the test had proven quite predictive in thses localities and that we now wanted to see how well it worked in a school district in this locale with this particular socio-economic status. (2) The other major purpose was to

assess the difference between small and large group testing. Hence, one testing would be done with small groups of four to six students while another would be completed for the class as a unit. Teachers were told they would be advised of test results before the small groups were tested. (3) The two accompanying tests were included in order to round out information on each child. The coding test was intended to measure the amount of "effort" the child was putting into the task. The self-esteem inventory was intended to measure how the children felt about themselves, their school and school work. It was stated that these two tests were necessary items in conducting test development.

At this point, an explanation sheet was given to the teachers which reiterated and elaborated on the above points (see Appendix H). In addition, they were given a set of pre-test material (see Appendix B1, B3) and the teacher's pre-experimental expectation measure (see Appendix E).

The teachers listened attentively and quickly looked over the materials given to them. They seemed to accept the above explanation and moved on to ask questions which would clarify their role in the research. The teachers expressed concern about public relations. They noted that parent-teacher conferences were scheduled for the late fall and that parents would want to know how their children scored on the various tests. The principal suggested that parents be told it was a group type research project and that individual assessment was not its main focus. In addition, it was decided to send a

letter of explanation to the parents of the participating children as a courtesy, informing them of the research (see Appendix I).

The teachers suggested they prepare students for testing by telling them it was part of a research program and had no implications for reading or math grouping or placement in next year's classes. In this way, it was hoped to reduce test anxiety in a school already low in test consciousness. The teachers then took out their plan books and a schedule was drawn up for pre-testing. At no time during this meeting was there the slightest indication that the teachers doubted the explanation presented by the E.

Steps 6 and 7: Pre-Testing

On September 21, 1970, the teachers returned to E the completed teacher pre-experimental expectation measure. It was important that the teachers complete these forms before pre-testing began.

The E was aided in executing the pre- and post-tests by an accomplice. She was a middle-aged woman who had taught in the school district for a number of years. Currently she was retired but maintained informal ties with the school. She was well known to the principal and other supporting staff (school nurse, librarian, etc.) but was not known to the four participating teachers. Her cooperation enhanced the credibility of the experiment since she was a trusted and respected confidante of the school.

Considerable thought was given in weighing the relative advantages and disadvantages of the accomplice or classroom teacher administering the pre-test. The decision to use the accomplice stemmed from several overarching considerations: (1) The teachers and principal advised that the teachers were not accustomed to testing. Hence, considerable variation in pre-test scores could occur from one class to another largely on the basis of variations in teacher test style. We wanted to standardize the pre-testing as much as possible to eliminate such "noise." (2) The teachers were reluctant to attempt pre-testing themselves because of their inexperience. It was felt that this request should be honored in order to ensure their continued cooperation. (3) Pre-testing by the accomplice increased the credibility of the experiment in the eyes of the teachers. Her presence emphasized the importance of the research and the commitment of the school to its competent execution.

Pre-testing in the four classrooms was completed on September 21-23, 1970. The tests were administered by the accomplice with the aid of the classroom teacher. The testing proceeded as follows: The accomplice entered the classroom at the appointed time. The teacher terminated the activity the children were engaged in and introduced the accomplice. The first test, which was handed out by the teacher, included two fact sheets. One was an explanation of the research-testing and the children's role in it

(see Appendix J). The other was the students' pre-experimental expectation measure (see Appendix F). The accomplice read aloud these fact sheets as well as the remaining tests as the students followed along. Students were seated in small groups of five or six each around a circular table. The teacher and accomplice circulated around the room to check that students were on the correct page and that they understood directions. Sufficient time was allowed for students to finish an item before the next one was read. Each of the three pre-tests was handed out and collected separately. If a student had a question about an item, the entire item was re-read. At the termination of the testing, the students were reminded that the tests would be re-administered again in a few weeks and that by then their scores would be available.

Step 8: Post-Testing:

Post testing took place over a ten day period extending from October 14 to October 27, 1970. A schedule was drawn up by E and the school principal beforehand, which took into account the needs of the research as well as the convenience of the students and teachers. An attempt was made to equalize the occurrence of any particular experimental treatment over the course of the ten days so that there would be no noticeable clustering. Post-testing occurred principally in the morning. A schema of the post-testing schedule appears in Tables 3 and 4.

These tables illustrate several counterbalancing techniques that were used in this design: 1) Experimental treatments were counterbalanced across classrooms. For example, Teacher 1 post-tested his small groups in the order 4, 3, 5, 1, 2, whereas Teacher 2 post-tested hers in the order 5, 4, 2, 3, 1. 2) Within a classroom, the order in which the battery was administered was counterbalanced. For example, Teacher 1 administered the post-tests on October 14 in the order: Iowa-Coding - Self Esteem Inventory. On October 15, the order was: Self Esteem Inventory - Iowa - Coding. These counterbalancing techniques were necessary to control for sequential effects.

The post-testing proceeded as follows: the accomplice and E entered the teacher's classroom at the appointed time. The teacher called the students who were to be tested that day from a list previously given to her. The small groups of five to six students then left the regular classroom and went with the E to a small room in another part of the school. There they were seated around a circular table. The E told them that their classroom teacher would arrive a few minutes later to administer again the battery of tests they took the previous month. The E stated that their performance on the first round of testing was available and that he thought that they would want to know how they did. The E then passed out a "report form" to each student which contained positive, average, or negative feedback (see Appendix K). The E read the contents of the report form aloud to the group. This was the manner in which the experimental induction for the students was

accomplished. He then asked each student to answer the question at the bottom of the form: "Do you think that's true about yourself?" Students circled either Yes or No and then the E collected the students' "report forms." In this way the success of the experimental induction was checked on.

Simultaneously, the experimental induction was being administered to the teacher. The accomplice gave her a similar "report form" which served to induce positive, average or negative expectations (see Appendix L). A single form was provided for each child and at the bottom the success of the induction procedure was checked upon by asking the teacher: "Do you accept this statement as reasonable?" The teacher circled Yes or No and the accomplice collected the forms. The accomplice remained in the classroom during the post-testing while the teacher left and went to the small room to join the small group for the post-testing. By the time she arrived the students were settled and ready to begin.

As the teacher entered the room the E told her the order of post-test administration for that day. He then seated himself nearby and activated the tape recorder. The teacher immediately began the post-testing and passed out and collected the test materials. Usually she seated herself on a desk overlooking the small group so that she could clearly observe their responses and behavior. From time to time she would walk around the group taking a closer look at the tests. This small group atmosphere maximized communication between teacher and student. The teacher read aloud all instructions and test

items as the children followed along. She waited for the group to finish an item before going on to the next one. The role of the E was simply to be custodian of the tape recorder. He allowed the teacher complete freedom to take as long as she thought necessary to complete the testing. If he were asked a question by her (e/g/. "Should I be going faster?") he answered by saying that she should do whatever she thinks is right.

The five to six students post-tested together all had the same induced expectations. This was necessary since the teacher was reading aloud the test items to the five to six students as a group. Homogeneous expectations for each sub-group was necessary since the teacher was communicating expectations for the entire sub-group of five to six students rather than post-testing each child individually.

In all, there were a total of 20 post-testing sessions (i.e., five treatments for four teachers). All of them were conducted in the same room thus providing a standard stimulus environment. The E was present for all post-testing sessions and he never observed a single instance of overt coaching on the part of the teacher. It was my observation that the teachers conducted these sessions in a professional manner. There was very little extraneous conversation during the testing as both the teachers and students seemed to be task-oriented.

Step 9: De-Briefing of School Staff

On November 20, 1970, the E met individually with the four participating teachers to de-brief them and to give them the actual test scores. Each of them was asked to re-state the purpose of the experiment to the E and then they were asked if they had developed any suspicions about the true intent of the experiment. Teacher 1 said he never suspected anything other than what he was told by the E. He had heard of the "Rosenthal effect" operating over longer periods of time but never suspected the entent of this experiment. Teachers 2 and 3 had never raised suspicions and neither had they heard about expectancy research. Teacher 4, however, did develop a suspicious attitude toward the beginning of the experiment. She could not accept the feedback given her for a particular child since it was so discrepant with her own view of him. She then developed her own hypothesis about the true intent of the experiment. She decided that it was a test to see how well teachers could predict students' performances. She pointed out that I had asked her to state her opinion about the children (i.e., use of teacher pre-experimental expectation measure) and then the tests were soon thereafter administered. She had never suspected, however, the true intent of the experiment. Teachers have been included in the data analysis as a factor, so the effect she exerted on the overall results can be determined.

Teachers were given the actual scores on the Iowa Test of Basic Skills and the Self-Esteem Inventory. They used these

as a learning experience both for themselves and the students and interpreted the scores to each child. In this way, the students were de-briefed.

In sum, the teachers accepted the experiment as a learning experience, and handled themselves in a professional manner. The teachers requested additional test scores and these were sent to them during the winter of 1970-1971.

Table 1
Assignment of students to experimental groups matched on four variables (n=100)

	Teacher 1	Teacher 2
NEGATIVE-NEGATIVE	Number = 4 Sex = 2F; 2M Iowa Mean = 38.7 Teachers pre = 7.0 Students pre = 2.2	Number = 5 Sex = 2M; 3F Iowa Mean = 44.6 Teachers pre = 5.6 Students pre = 1.4
POSITIVE-POSITIVE	Number = 4 Sex = 2F; 2M Iowa Mean = 39.0 Teachers pre = 7.2 Students pre = 1.7	Number = 6 Sex = 3F; 3M Iowa Mean = 45.0 Teachers pre = 6.2 Students pre = 1.8
AVERAGE-AVERAGE	Number = 4 Sex = 2F; 2M Iowa Mean = 39.0 Teachers pre = 7.2 Students pre = 1.7	Number = 6 Sex = 3F; 3M Iowa Mean = 44.6 Teachers pre = 6.0 Students pre = 1.3
POSITIVE-NEGATIVE	Number = 5 Sex = 2F; 3M Iowa Mean = 39.0 Teachers pre = 7.8 Students pre = 2.2	Number = 6 Sex = 3F; 3M Iowa Mean = 45.3 Teachers pre = 6.4 Students pre = 1.5
NEGATIVE-POSITIVE	Number = 4 Sex = 2F; 2M Iowa Mean = 39.0 Teachers pre = 7.0 Students pre = 2.0	Number = 5 Sex = 2F; 3M Iowa Mean = 44.8 Teachers pre = 6.3 Students pre = 1.8

	Teacher 3	Teacher 4
NEGATIVE-NEGATIVE	Number = 5 Sex = 3F; 2M Iowa Mean = 39.4 Teachers pre = 6.2 Students pre = 1.4	Number = 5 Sex = 2F; 3M Iowa Mean = 38.6 Teachers pre = 6.3 Students pre = 2.0
POSITIVE-POSITIVE	Number = 5 Sex = 2F; 3M Iowa Mean = 39.6 Teachers pre = 6.0 Students pre = 1.2	Number = 5 Sex = 2F; 3M Iowa Mean = 38.4 Teachers pre = 7.5 Students pre = 2.0
AVERAGE-AVERAGE	Number = 6 Sex = 3F; 3M Iowa Mean = 39.5 Teachers pre = 6.5 Students pre = 1.1	Number = 5 Sex = 1F; 4M Iowa Mean = 38.6 Teachers pre = 1.8 Students pre = 1.8
POSITIVE-NEGATIVE	Number = 5 Sex = 2F; 3M Iowa Mean = 39.4 Teachers pre = 5.8 Students pre = 1.2	Number = 5 Sex = 2F; 3M Iowa Mean = 38.6 Teachers pre = 6.6 Students pre = 1.8
NEGATIVE-POSITIVE	Number = 5 Sex = 3F; 2M Iowa Mean = 39.2 Teachers pre = 5.8 Students pre = 1.4	Number = 5 Sex = 2F; 3M Iowa Mean = 38.4 Teachers pre = 6.6 Students pre = 1.8

Table 2
Steps In Execution of Experiment

1.) May 1970
School Board approves research.

2.) September 10, 1970
Meeting with school principal.

3.) September 14, 1970
Memorandum sent to participating teachers by principal advising them of research plans.

4.) September 17, 1970
Meeting with four participating teachers and principal to discuss details of the research.

5.) September 19, 1970
Letters of explanation mailed to parents of participating children.

6.) September 21, 1970
Teachers complete and return pre-experimental expectation measure.

7.) September 21-23, 1970
Pre-test administered.

8.) October 14-27, 1970
Post-test administered.

9.) November 20, 1970
Teachers and students de-briefed and test results given to teachers.

Table 3

SCHEMATIC DIAGRAM OF SCHEDULE FOR POST-TESTING

	Oct. 14	Oct. 15	Oct. 16	Oct. 19	Oct.20
9:15 AM	Teacher 2 Gp: Av-Av Order: Iowa Self Coding	Teacher 3 Gp: Pos-Pos Order: Iowa Coding Self	Teacher 3 Gp: Neg-Neg Order: Self Iowa Coding	Teacher 2 Gp: Neg-Pos Order: Coding Iowa Self	Teacher 4 Gp: Pos-Pos Order: Coding Iowa Self
10:35 AM	Teacher 1 Gp: Neg-Pos Order: Iowa Coding Self	Teacher 1 Gp: Pos-Neg Order: Self Iowa Coding	Teacher 4 Gp: Neg-Neg Order: Iowa Self Coding	Teacher 1 Gp: Av-Av Order: Coding Self Iowa	Teacher 3 Gp: Neg-Pos Order: Coding Self Iowa
12:45PM					

	Oct. 21	Oct. 22	Oct. 23	Oct. 26	Oct. 27
9:15AM	Teacher 2 Gp: Pos-Pos Order: self Coding Iowa	Teacher 3 Gp: Av-Av Order: Iowa Coding Self	--	Teacher 2 Gp: Neg-Neg Order: Coding Iowa Self	Teacher 4 Gp: Neg-Pos Order: Coding Iowa
10:35 AM	Teacher 1 Gp: Neg-Neg Order: Iowa Coding Self	Teacher 1 Gp: Pos-Pos Order: Self Iowa Coding	Teacher 2 Gp: Pos-Neg Order: Iowa Self Coding	Teacher 4 Gp: Av-Av Order: Iowa Self Coding	Teacher 3 Gp: Pos-Neg Order: Self Iowa Coding
12:45 PM		Teacher 1 Gp: Pos-Neg Order: Self Coding Iowa			

Table 4

ORDER OF EXPERIMENTAL GROUP TESTING BY TEACHER

Teacher 1	Teacher 2	Teacher 3	Teacher 4
4	5	2	1
3	4	1	2
5	2	4	3
1	3	5	4
2	1	3	5

KEY:
1 = NEGATIVE-NEGATIVE EXPERIMENTAL GROUP
2 = POSITIVE-POSITIVE EXPERIMENTAL GROUP
3 = POSITIVE-NEGATIVE EXPERIMENTAL GROUP
4 = NEGATIVE-POSITIVE EXPERIMENTAL GROUP
5 = AVERAGE-AVERAGE EXPERIMENTAL GROUP

APPENDIX N1

VOICE RATINGS

Below you will find a list of three attributes that are commonly used to describe voices. I would like you to rate them on each of these attributes by circling one of the numbers from 1 to 10.

Please work through the ratings as quickly as possible. Trust your first impression and do not mull over your ratings. If you do not quite understand one of the attributes or voices, interpret it in a way that makes sense to you.

You will hear 20 seconds of voice followed by a pause during which you are to rate the voice. Rest periods will be interspursed throughout the ratings. In all, there are 60 voices to be rated.

Remember that we are interested in your spontaneous reaction, in your first impression of the voice. Somewhat contradictory ratings within a rating list happen quite frequently. Please do not attempt to avoid these contradictions by checking prior ratings. Record your first judgment for each attribute right away and do not change it unless you make an error.

APPENDIX N2

ATTRIBUTES

Name of Rater:_____ Voice Number_____

No. 1

1 2 3 4 5 6 7 8 9 10
PLEASANT UNPLEASANT

No. 2

1 2 3 4 5 6 7 8 9 10
WARM COLD

No. 3

1 2 3 4 5 6 7 8 9 10
PROFESSIONAL PERSONAL
BUSINESS-LIKE

Appendix O

Table 1

Intercorrelation matrix for Negative-Negative Group (N=19)

	Teachers Pre	Students Pre	SEI Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.22	-0.60	.49	.83	40.42	6.45
Teachers Pre-Expectations		-0.22	.12	.22	5.00	3.05
Students Pre-Expectations			-0.68	-0.70	1.73	.73
SEI Pre-Test				.39	65.05	19.62
Iowa Post-Test					41.05	8.91

Appendix O

Table 2

Intercorrelation matrix for Positive-Positive Group (N=20)

	Teachers Pre	Students Pre	SEI Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.38	-0.15	.07	.81	40.80	7.09
Teachers Pre-Expectations		-0.34	-0.40	.29	4.65	3.46
Students Pre-Expectations			-0.27	-0.18	1.70	.57
SEI Pre-Test				-0.01	61.40	15.28
Iowa Post-Test					40.05	8.21

Appendix O

Table 3

Intercorrelation matrix for Negative-Positive Group (N=19)

	Teachers Pre	Students Pre	SEI Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.38	.44	0.29	.77	40.42	6.30
Teachers Pre-Expectations		.30	.03	.37	5.05	3.29
Student Pre-Expectations			-0.29	.32	1.73	.56
SEI Pre-Test				-0.23	65.15	13.50
Iowa Post-Test					40.89	8.29

Appendix O

Table 4

Intercorrelation matrix for Positive-Negative Group (N=21)

	Teachers Pre	Students Pre	SEI Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.51	-0.45	.30	.85	40.90	7.42
Teachers Pre-Expectations		-0.21	.15	.54	5.71	3.16
Students Pre-Expectations			-0.55	-0.45	1.66	.65
SEI Pre-Test				.26	66.28	14.83
Iowa Post-Test					40.76	8.34

Appendix O

Table 5

Intercorrelation matrix for Average-Average Group Group (N=21)

	Teachers Pre	Students Pre	SEI Pre-Test	Iowa Post	Mean	S.D.
Iowa Pre-Test	.53	-0.02	.55	.76	40.66	6.93
Teachers Pre-Expectations		.02	.44	.37	5.00	3.40
Students Pre-Expectations			-0.33	-0.13	1.47	.51
SEI Pre-Test				.55	70.00	12.74
Iowa Post-Test					43.95	7.01

Appendix O

Table 6

Repeated Measures Analysis of Variance of Iowa Post-Test Items. Treatment Groups (5) X First/Second Half (2). Fourth Grades

	d.f.	Mean Square	F
TREATMENT GROUPS	4	8.463	.73[a]
BETWEEN SUBJECTS	67	11.459	
FIRST/SECOND HALF	1	0.811	.10
FIRST/SECOND HALF X TREATMENT	4	4.612	.56
FIRST/SECOND HALF X SUBJECTS WITHIN GROUPS	67	8.151	

a. P= .60

Appendix O

Table 7

Repeated Measures Analysis of Variance of Iowa Post-Test Items. Treatment Groups (5) X First/Second Half (2). Fifth Grade

	d.f.	Mean Square	F
TREATMENT GROUPS	4	17.310	1.56[a]
BETWEEN SUBJECTS	23	11.05	
FIRST/SECOND HALF	1	15.402	2.05[b]
FIRST/SECOND HALF X TREATMENT	4	5.574	0.74
FIRST/SECOND HALF X SUBJECTS WITHIN GROUPS	23	7.48	

a. P= .21

b. P $<$ .10

Appendix O

Table 8

Analysis of Variance on Pleasant-Unpleasant Dimension. 3 (Beginning-Middle-End) X 5 (Treatment Groups)

	d.f.	Mean Square	F
Beginning-Middle-End	2	42.62	1.94[a]
Treatment Group	4	24.91	1.13
Beginning-Middle-End X Treatment Groups	8	10.44	.47
ANOVA Error	44	21.91	

a. P= .15

Appendix O

Table 9

Analysis of Variance of Warm-Cold Dimension. 3 (Beginning-Middle-End) X 5 (Treatment Groups)

	d.f.	Mean Square	F
Beginning-Middle-End	2	41.42	2.51[a]
Treatment Groups	4	12.94	.78
Beginning-Middle-End X Treatment Groups	8	18.42	1.11
ANOVA Error	44	16.49	

a. P= .09

Appendix O

Table 10

Analysis of Co-Variance for Self-Esteem Inventory. 5 (Treatment Groups) x 4 (Sections)

	d.f.	Mean Square	F
Treatment Groups	4	82.40	.94
Sections	3	31.03	.35
Treatment Groups X Sections	12	128.64	1.47[a]
Regression	1	14946.13	170.96
Adjusted Error	79	87.42	
ANOVA Error	80	273.15	

a/ P= .15

Appendix O

Table 11

Analysis of Co-Variance for Sub-Scales on Self-Esteem Inventory

A. General Sub-Scales:	T Group F = 1.32 Sections F = 0.64 Interaction F = 1.17
B. Social Sub-Scale:	T Group F = 0.73 Sections F = 0.28 Interaction F = 1.81
C. Home Sub-Scale:	T Group F = 1.18 Sections F = 0.60 Interaction F = 0.33
D. School Sub-Scale:	T Group F = 1.00 Sections F = 0.97 Interaction F = 1.37
E. Lie Sub-Scale	T Group F = 2.04 Sections F = 1.16 Interaction F = 0.80

All relationships were non-significant tested at the .05 level.

Appendix O

Table 12

Analysis of Co-variance for Coding Test. 5 (Treatment Groups) X 4 (Sections)

	d.f.	Mean Square	F
Treatment Groups	4	22.68	.67
Sections	3	82.35	2.46[a]
Treatment Groups X Sections	12	62.97	1.88[b]
Regression	1	2766.12	82.77
Adjusted Error	79	33.41	
ANOVA Error	80	67.57	

a. P= .06
b. P= .04

Appendix O

Table 13

Mean Score for Subjects Above and Below the Mean on Iowa Test of Basic Skills and Self-Esteem Inventory Pre and Post Test for Negative-Negative Group.

	Iowa Test		Self-Esteem Inventory	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Above Mean	56.36	56.09	78.00	79.00
Below Mean	41.00	40.50	50.66	51.77

Appendix O

Table 14

Mean Score for Subjects Above and Below the Mean on Iowa Test of Basic Skills and Self-Esteem Inventory Pre and Post Test for Positive-Positive Group.

	Iowa Test		Self-Esteem Inventory	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Above Mean	57.18	55.18	71.27	68.18
Below Mean	41.33	39.77	49.33	52.88

Appendix O

Table 15

Mean Score for Subjects Above and Below the Mean on Iowa Test of Basic Skills and Self-Esteem Inventory Pre and Post Test for Positive-Negative Group.

	Iowa Test		Self-Esteem Inventory	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Above Mean	61.00	60.00	84.28	84.00
Below Mean	44.92	43.42	57.28	60.28

Appendix O

Table 16

Mean Score for Subjects Above and Below the Mean on Iowa Test of Basic Skills and Self-Esteem Inventory Pre and Post Test for Negative-Positive Group.

	Iowa Test		Self-Esteem Inventory	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Above Mean	56.80	54.60	78.75	83.75
Below Mean	41.77	43.77	55.27	60.36

Appendix O

Table 17

Mean Score for Subjects Above and Below the Mean on Iowa Test of Basic Skills and Self-Esteem Inventory Pre and Post Test for Average-Average Group.

	Iowa Test		Self-Esteem Inventory	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Above Mean	58.54	59.00	79.33	79.83
Below Mean	40.40	48.00	57.55	57.55

XI. BIBLIOGRAPHY

Adair, J.G. Demand characteristics or conformity? Suspiciousness of deception and experimenter bias in conformity research. Unpublished Manuscript, University of Manitoba, 1968.

Adair, J.G. & Epstein, J. Verbal cues in the mediation of experimenter bias. Paper presented at the Meeting of the Midwestern Psychological Association, Chicago, May, 1967.

Anderson, D.F. & Rosenthal, R. Some effects of interpersonal expectancy and social interaction on institutionalized retarded children. Proceedings of the Seventh-Sixth Annual Convention of the American Psychological Association, 1968, 479-480.

Argyris, Chris. The incompleteness of social-psychological theory. American Psychologist, 1969, 24, No. 10, 893-908.

Barber, Theodore X. Invalid arguments, postmortem analyses, and the experimenter bias effect. Journal of Counseling and Clinical Psychology, 1969, 33, (1), 11-14.

Barber, Theodore X., Forgione, Albert, Chaves, John, Calverley, Davis S., Peake, John D. & Bowen, Barbara. Five attempts to replicate the experimenter bias effect. Journal of Counseling and Clinical Psychology, 1969, 33, (1), 1-6.

Barber, Theodore X. & Silver, Maurice. Pitfalls in data analysis and interpretation: A reply to Rosenthal. Psychological Bulletin Monograph, 1968, 70, 48-62.

Barber, Theodore X.& Silver, Maurice. Fact, fiction and the experimenter bias effect. Psychological Bulletin Monograph, 1968, 70, 1-29.

Beez, W.V. Influence of biased psychological reports on teacher behavior and pupil performance. Proceedings of the Seventy-Sixth Annual Convention of the American Psychological Association, 1968, 605-606.

Biegen, D.A. Unpublished data. University of Cincinnati, 1968.

Bootzin, R.R. The experimenter: a credibility gap in psychology. Unpublished Manuscript, Purdue University, 1968.

Brager, Gary L. A study of the phenomenon of experimenter bias using three types of classroom tasks. Dissertation Abstracts, 1969, 29, (12-A), 4316-4317.

Burnham, J.R. & Hartsough, D.M. Effect of experimenter's expectancies on children's ability to learn to swim. Paper Presented at the Meeting of the Midwest Psychological Association, Chicago, May, 1968.

Buros, Oscar K. Tests in Print. New Jersey: The Gryphon Press, 1961.

Canavan, Donnah. Development of individual differences in the perception of value and risking-taking styles. Ph.D. Dissertation. New York: Teachers College, Columbia University (in preperation), 1968.

Carlson, J.A. & Hergenhahn, B.R. Use of tape-recorded instructions and a visual screen to reduce experimenter bias. Unpublished Manuscript, Hamline University, 1968.

Clairborn, W.L. An investigation of the relationship between teacher expectancy, teacher behavior, and pupil performance. Unpublished Doctoral Dissertation, Syracuse University, 1968.

Clark, Kenneth B. Dark Ghetto: Dilemmas of Social Power. New York: Harper & Row, 1965.

Coleman, James S., Campbell, Ernest O., McPartland, James, Halson, Carol, Mood, Alexander M., Weigelf, Frederic D., & York, Robert L. Equality of Educational Opportunity. Washington, D.C.: United States Department of Health, Education and Welfare, 1966.

Conn, L.K., Edwards, C.N., Rosenthal, R. & Crowne, D. Perception of emotion and response to teacher's expectancy by elementary school children. Psychological Reports, 1968, 22, 27-34.

Coopersmith, Stanley. The Antecedents of Self-Esteem. San Francisco: W.H. Freeman & Company, 1967.

Duncan, S. & Rosenthal, R. Vocal emphasis in experimenters' instruction reading as unintended determinant of subjects' responses. Language and Speech, 1968, 2, Part 1, 20-26.

Epps, Edgar G. Negro achievement motivation and performance: An overview. Journal of Social Issues, Vol. 25, No. 3, Summer 1969, 5-11.

Fisher, S., Cole, J.O., Rickels, K., & Uhlenhuth, E.H. Drug-set interaction: The effect of expectation on drug response in out-patients. In P. B. Bradley, Flugel, F. & Hoch, P. (Eds.), Neuro-Psychopharmacology, 3, Elsevier, New York, 1964, 149-156.

Flanagan, J.C. Test of General Ability: Technical Report. Science Research Associates, Chicago, 1960.

Flowers, C.E. Effects of arbitrary accelerated group placement on the testing of academic achievement of educationally disadvantaged students. Unpublished Doctoral Dissertation, Teachers College, Columbia University, 1966.

Friedman, H. Magnitude of experimental effects and a table for its rapid estimation. Psychological Bulletin, 1968 70, 245-251.

Fromm-Reichmann, Freida. Principles of Intensive Psychotherapy. University of Chicago Press, Chicago, Ill, 1950.

Goldstein, A.P. Therapist-Patient Expectancies in Psychotherapy. New York: Pegramon Press, 1962,

Gumpert, Peter & Gumpert, Carol. On the psychology of expectation in the classroom. The Urban Review, 1968, 3, No. 1, 21-22.

Haskett, Mary S. An investigation of the relationship between teacher expectancy and pupil achievement in the special education class. Dissertation Abstracts, 1969, 29, (12-A), 4348-4349.

Heiserman, M.S. The relationship between teacher expectations and pupil occupational aspirations. Unpublished Master's Thesis, Iowa State University, Ames, 1967.

Homme, Lloyd E., DeBaca, P., Devine, J.V., Steinhorst, R.& Rickert, E.J. Use of the premack principle in controlling the behavior of nursery school children. Journal Experimental Analysis of Behavior, 1963, 6, 544.

Homme, Lloyd E. Human motivation and environment. Paper Read at the University of Kansas Symposium, March, 1965.

Honig, W.K. (Ed.) Operant Behavior: Areas of Research and Application. New York: Appleton-Century-Crofts, 1966.

Hovland, Carl I., Lumsdaine, Arthur, Sheffield, Fred D. Experiments On Mass Communication. New Jersey: Princeton University Press, 1949, 3, 302-305.

Jacob, Theodore. The mediation of the experimenter bias effect as a function of "demand characteristics", experimenter "investment" and the nature of the experimental task. Dissertation Abstracts International, 1970, 30, (7-B), 3372-3373.

Johnson, R.W. Subject performance as affected by experimenter expectancy, sex of the experimenter, and verbal reinforcement. Unpublished Masters' Thesis, University of New Brunswick, 1967.

Kennedy, John J. Experimenter outcome bias in verbal conditioning: a failure to detect the Rosenthal effect. Psychological Reports, 1969, 25, (2), 495-500.

Kessell, Paul & Barber, K.J. Jr. Experimenter-subject interaction in verbal conditioning: a review of the literature. Psychological Reports, 1968, 22, (2), 59-74.

Kester, Scott W. The communication of teacher expectation and their effects on the achievement and attitude of secondary school pupils. Dissertation Abstracts International, 1969, 30, (4A), 1434-1435.

Levy, Leon H. Reflections on replications and the experimenter bias effect. Journal of Counseling and Clinical Psychology, 1969, 33, (1), 15-17.

Lindquist, E.F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton, Mifflin Company, 1953.

Lindquist, E.F. Test Manual for Iowa Test of Basic Skills. Boston: Houghton, Mifflin Company, 1964.

McFall, Richard M. & Schenkein, Diane. Experimenter expectancy effects, need for achievement and field dependence. Journal of Experimental Research in Personality, 1970, 4, (2), 122-128.

Masling, J. Differential indoctrination of examiners and rorschach responses. Journal of Counseling Psychology, 1965, 29, 198-201.

McClelland, D.C., Atkinson, J.W., Clark, R.A. & Lowell, E.L. The Achievement Motive. New York: Appleton-Century-Croft, 1953.

Meichenbaum, D.H., Bowers, K.S. & Ross, R.R. A behavioral analysis of teacher expectancy effect. Unpublished Manuscript, University of Waterloo, 1968.

Merton, Robert K. The Self-fulfilling prophecy. Antioch Review, 1948, 8, 193-210.

Milmore, Susan, Novey, M.S., Kagan, J. & Rosenthal, Robert. The mother's voice: postictor of aspects of her baby's behavior. Proceedings, Seventy-Sixth Annual Convention, American Psychological Association, 1968.

Moffatt, M.C. Unpublished Data. University of British Columbia, 1966.

Morgan & King. Introductory Psychology. New York: McGraw Hill, 1971.

Pitt, C.C.V. An experimental study of the effects of teacher's knowledge or incorrect knowledge of pupil I.Q.s on teacher's attitudes and practices and pupil's attitudes and achievement. Unpublished Doctoral Dissertation, Columbia University, 1956.

Rosenthal, Robert. Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. Psychological Bulletin Monograph, 70, 1968, 30-47.

Rosenthal, Robert. On not so replicated experiments and not so null results. Journal of Counseling and Clinical Psychology, 1969, 33, (1), 7-10.

Rosenthal, Robert. Unintended effects of the clinician in clinical interaction: A taxanomy and a review of clinical expectancy effects. Australian Journal of Personality, 1969, 21, (1), 1-20.

Rosenthal, Robert. Experimenter effects. In Rosnow, R.L. & Rosenthal, Robert (Eds.). Artifacts in Social Research. New York: Academic Press, 1971.

Rosenthal, Robert L. Empirical vs. decreed validation of clocks and tests. Unpublished rebuttal to Robert L. Thorndike's review of Pygmalion In The Classroom, 1969.

Rosenthal, Robert & Anderson, D.F. Teacher behavior and the mediation of teacher expectancy effects. Unpublished Data, Harvard University, 1969.

Rosenthal, Robert & Evans, J. Unpublished Data, Harvard University, 1968.

Rosenthal, Robert & Jacobson, Leona. Pygmalion In The Classroom: Teacher Expectation and Pupil's Intelligence Development. New York: Holt, Rinehart and Winston, 1968.

Rosenthal, Robert & Fode, K.L. Three experiments in experimenter bias. Psychological Reports, 12, 1963, 491-511.

Rosenthal, Robert. Teacher expectation and pupil learning. Paper Presented for a Conference on The Unstudied Curriculum, sponsored by the Association for Supervision and Curriculum Development, Washington, D.C., January, 1969.

Rosenthal, Robert. Experimenter Effects In Behavioral Research. New York: Appleton-Century-Crofts, 1966.

Scheurer, Klaus. Personal Communication, 1970.

Shapiro, A.K. A contribution to a history of the placebo effect. Behavioral Science, 5, 1960, 109-135.

Sheehan, Peter. E-expectancy and the role of awareness in verbal conditioning. Psychological Reports, 1969, 24, (1), 203-206.

Silverman, I. The effects of experimenter outcome expectancy on latency of word association. Journal of Clinical Psychology, 24, 1968, 60-63.

Skinner, B.F. The Technology of Teaching. New York: Appleton-Century-Crofts, 1968.

Skinner, B.F. Science and Human Behavior. New York: The Macmillan Company, 1953.

Snow, Richard E. Unfinished Pygmalion - Review of Pygmalion in the Classroom by Robert Rosenthal. Contemporary Psychology, 14, 1969, 197-199.

Stevenson, H.W. & Odom, R.D. Visual reinforcement with children. Unpublished Manuscript, University of Minnesota, 1963.

Thorndike, Robert L. Review of Pygmalion in the Classroom. American Educational Research Journal, 5, (4), 1968, 708-711.

Troffer, S.A. & Tart, C.T. Experimenter bias in hypnotism performance. Science, 145, 1964, 1330-1331.

Weick, K.E. When prophecy fails: The fate of dissonance theory. Psychological Reports, 16, 1965, 1261-1275.

Weiner, Bernard & Kukla, Andy. An attribution analysis of achievement motivation. 1969, in press.

Weiner, Bernard, Frieze, Irene, Kukla, Andy, Reed, Linda, Rest, Stanley & Rosenbaum,Robert M. An attribution (cognitive) model of motivation. Unpublished Manuscript, University of California, 1969.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw Hill Book Co., Inc., 1962.

Witkin, H.A., Lewis, H.B., Hertzman, M., Machover, K., Meissner, P. & Wapner, S. Personality through Perception: An experimental and Clinical Study. New York: Harper, 1954.

Wyatt, D.F. & Campbell, Don T. A study of interviewer bias as related to interviewers' expectations and own opinions. International Journal of Opinion and Attitude Research, 4, 1950, 77-83.

Zoble, E.J. Interaction of subject and experimenter expectancy effects in a tone length discrimination task. Unpublished Thesis, Franklin and Marshall College, 1968.