70-24,479

MASE, Howard Stuart, 1935-

The City University of New York, Ph.D., 1970
Psychology, industrial

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE EFFECTS OF INTERPERSONAL EXPECTANCY UPON LEARNING ACHIEVEMENT IN AN INDUSTRIAL SETTING

by

Howard Mase

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

1970

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 15, 1970
date

Chairman of Examining Committee

5-19-70
date

Executive Officer

Supervisory Committee

The City University of New York

ACKNOWLEDGEMENTS

I would first like to express my deep gratitude to Dr. Leonard Simon, Chairman of my committee, for his guidance and support throughout the study. I am also grateful to the other members of that committee, Dr. Ross Harrison and Dr. Harold Wilensky, for their constructive criticism and suggestions.

There are many people at First National City Bank to whom I owe thanks for their assistance and encouragement. Dr. Norman Willard, for his willingness to sponsor the research; Eileen Bresnihan, for typing the manuscript; David Franzen and Amy Tenzer, for their help with the statistical analyses; Dan McElwreath and Jack Starr for their comments and suggestions.

Finally, I want to express my deep appreciation to my wife, Roberta Mase, for her moral support and tolerance throughout these years of graduate study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FORMULATION OF THE PROBLEM

We all have experienced the satisfying feeling that comes with having successfully predicted another person's behavior in a given situation. For the most part, accuracy in predicting can be traced to knowledge of the person's behavior in similar situations. There is now some evidence to indicate that another variable may contribute heavily to our accuracy in making interpersonal predictions and that variable is the prediction or the prophecy itself. The essence of this concept is that one's prophecy about another leads to a behavioral treatment of that person which tends to fulfill the prophecy - - thus creating the self-fulfilling prophecy. For example, when a teacher expects a student to do well, her behavior towards that student may be laden with smiles, encouragement or any other response that will enhance his chances to perform as expected. Conversely, if she expects him to do poorly, she may communicate that expectancy through avoidance or disapproving behavior thus increasing the chances that her prediction will prove correct.

The self-fulfilling prophecy is most significant in those situations where the inputs that form the expectancy are incorrect thus producing a false prophecy. If, in our example, the teacher expectancy had been based on erroneous, incomplete or even irrelevant information, her subsequent chain of interacting behavior with that student would have been ill-founded. If the prophecy prevails, the student will suffer the punishment or reap the rewards of a false

expectancy. A "good" student might do poorly and a "poor" student might do well. The latter situation is, of course, a desired one but it is the former that most often occurs in real-life. Consider the white teacher and his interaction with black students. Are his expectations for these students based on specific information or does a social class or racial bias form the basis for his expectancies? Consider the psychiatrist and his belief that the lower-class schizophenic has small chance of recovery. Is his expectancy based on hard evidence or do social-class differences lead him to that prediction? If it is indeed a real phenomenon, then clearly the self-fulfilling prophecy plays a major role in the behaviors and outcomes of interpersonal relations. There is sufficient evidence in the literature to suggest that it is real - - evidence from the laboratory and more important, evidence from real-life. Before addressing these findings it would be appropriate, from an historical standpoint, to first mention Merton's contribution to the conceptual development of the self-fulfilling prophecy.

Although not the first to describe its manifestations, almost all reviews of the self-fulfilling prophecy begin with Merton's (1949) application of the concept to such large scale social and economic phenomena as race relations and bank failures. It is interesting to note that Merton's definition of the concept covered only those circumstances where a false definition of the situation evokes a new behavior which makes the originally false conception come

true. Merton appears to exclude from his definition any predictions or expectancies that are based on true or correct information. Under the broader concept of interpersonal expectancy effects, expectancies based on correct information would also be included. In view of Merton's definition, perhaps the self-fulfilling prophecy should be limited to a very specific interpersonal expectancy effect; one that is based on false information only.

Expectancy Effects in the Laboratory

The great majority of laboratory studies investigating interpersonal expectancy effects deals with the problem of experimenter-subject interaction. Rosenthal (1966) has presented a comprehensive review of the investigations that demonstrate the effects of experimenter expectancy upon experimental outcomes. Based on the assumption that all behavioral scientists have some expectation about the results of their studies, Rosenthal and others have attempted to show that these expectations may become self-fulfilling prophecies (i.e., subject performance is determined by the experimenter's prophecy).

The research program carried out by Rosenthal was designed to disentangle the self-fulfilling aspect of a prophecy from its reality or knowledge components. As indicated previously, accuracy in predicting another's behavior does not necessarily mean that the prophecy led to its own accuracy. If students who are expected to do poorly perform as expected the accuracy

of the prediction might very well be a consequence of knowledge of past performance rather than a self-fulfilling effect.

Most of the studies investigating experimenter-expectancy effects with human subjects used a person - perception task as the criterion measure (e.g., Rosenthal and Fode, 1963B; Friedman, Kurland and Rosenthal, 1965). In this task subject is shown a series of photographed faces by a subject-experimenter and is asked to rate these photographs as to the degree of success or failure shown in the face. The rating scale employed ran from minus 10 (extreme failure) to plus 10 (extreme success). The ten photographs were selected so that, on the average, they would be perceived as quite neutral, with an average rating score of zero. Half of the subject-experimenters were told to expect the photos to be seen as successful (ratings of plus 5) and the remaining half were led to believe that the photos would be rated as unsuccessful (ratings of minus 5).

The overall result of these experiments are currently in great dispute. Rosenthal (1969) claims that the data supports the self-fulfilling prophecy theory, namely, that experimenters generally tended to obtain the data they expected to obtain. Experimenters expecting positive ratings obtained positive ratings, and experimenters expecting negative ratings obtained negative ratings.

Because of the potential impact of these findings upon the results of all psychological research, this research area has been subject to extremely close review and analysis. A recent review of 31 studies of the experimenter expectancy effect was conducted by Barber and Silver (1968). Concentrating primarily upon the statistical analyses of these studies, these authors made a study-by-study examination of the results and concluded that in most of these studies (19 of 31) the experimenter expectancy effect was not clearly demonstrated. Pointing to several inadequacies in the analysis of data procedures, Barber and Silver develop a very strong argument for questioning the accuracy of the findings. They note, for example, that in several studies an overall statistical analysis was not performed to exclude chance findings.

Barber and Silver's conclusions are obviously relevant in a general way to all research on interpersonal expectancy effects. The question of the validity of interpersonal expectancy effects can only be clarified through additional research. In fairness, it should be noted that Barber and Silver's criticisms are based solely upon laboratory experiments concerned with experimenter-subject interaction and even in this group of studies, they admitted to twelve experiments which did show an expectancy effect.

Some of the strongest support for expectancy theory can be found in studies of experimenter-expectancy effect where the

subjects were animals (Rosenthal, 1969 reports nine such studies). Rosenthal and Fode (1963A) and Rosenthal and Lawson (1964), for example, conducted investigations of rat learning performance. In these studies, some of the subject-experimenters were told that their rats were "bright" and should show rapid learning, and some were advised that their animals were "dull" and should show little evidence of learning. Both experiments, in accord with the hypothesis, showed that "bright" rats indicated significantly better learning results than did "dull" rats ($P < .01$).

In a study of the number of contractions and head turns shown by flatworms (planaria) during a two-second exposure to light, Cordaro and Ison (1963) advised several students conducting the experiment that their worms were previously conditioned to light and would probably show a high response rate. Other students were told that their flatworms were not conditioned and that they should not expect too much. The results were significant and in the predicted direction - students expecting more turning and contracting reported more turning and contracting ($P < .001$).

Although Barber and Silver classify the three above-mentioned animal studies as having shown the experimenter expectancy effect, they also point out that in each of these studies the task was such that errors (intentional or unintentional)

in recording the criterion measures were probable. Judging what constitutes a turn or a contraction in a flatworm, for instance, is considered highly ambiguous and subject to individual interpretation. It is also possible that data may have been falsified by prodding or handling the rat to perform as expected or by out-and-out fabrication of results. Barber and Silver's argument disputes Rosenthal's contention that expectancy effects are the result of unintentional and subtle communication between subject and experimenter. While a meaningful consideration, their point does not invalidate the notion of the self-fulfilling prophecy; it simply takes issue with Rosenthal's view as to the process involved. It would seem at least as important (if not more so) to know the extent to which experimenters will go to fulfill their prophecy. If an experimenter consciously "cheats" to get the desired behavior or if a teacher is aware that she is behaving in a special way to make Johnny perform as she expects him to; then interpersonal expectations take on even greater significance as a factor in shaping behaviors.

The person-perception studies described previously are not the only laboratory studies of expectancy effects in which human subjects were used. Rosenthal (1969) summarizes a total of 33 investigations of experimenter expectancy effects in areas of: human learning and abilities (10 studies), psychophysical judgments (9), reaction time (3), inkblot tests (5) and structured laboratory interviews (6). A few examples of these studies are presented below.

Rosenthal (1969) reports Larrabee & Kleinsasser's (1967) unpublished study on the effects of expectancy upon performance on the Wechsler Intelligence Scale for Children. Five experimenters administered the test to 12 sixth-grade children of average intelligence. Each subject had the even numbered items administered by one experimenter and the odd numbered items by another experimenter. One of the two administrators was told that the child had above average intelligence and the other was told that the child was of below average intelligence. Results of the study showed that when the child's experimenter had a positive expectation (i.e., that the child had above average intelligence), the total IQ earned was over 7 points higher on the average than when the child's experimenter thought the child had low average intelligence. The average difference was over 10 points on the verbal subtest.

Silverman (1968) used 20 advanced psychology students to administer a word association test to over 300 students of introductory psychology. Half of the experimenters were led to believe that certain subjects would exhibit longer response latencies than would others. The remaining experimenters were not given any expectation. Results showed that when compared to the control group, experimenters who expected longer latencies obtained them (P .02).

Marwit and Marcia (1967) performed an investigation of the effects of tester expectancy upon Rorschach responses.

Utilizing 36 undergraduate students of experimental psychology as testers and 33 undergraduate students of introductory psychology as subjects, they were able to demonstrate the effects of expectancy upon the number of responses to the inkblots. Twenty-five of the experimenters expected their subjects to give many Rorschach responses and the remaining eleven were led to expect few Rorschach responses from their subjects. Results showed that subjects who were expected to give more responses did give significantly more responses than those subjects from whom a low number of responses were expected (P .05). Masling (1965) performed a similar study utilizing type of response (animal or human) and also obtained results that were significant in the predicted direction (P .04).

The studies described above showed some evidence in support of Rosenthal's theory regarding interpersonal expectancy effects. However, there have been several investigations that did not find evidence for the hypothesis. Rosenthal (1969) provides a summary table indicating the percentage of studies that showed expectancy effects at .10 and .05 levels of probability (page 65, draft version)

Research Area	Number of Studies	Percentage of Studies at $p \leq .10$	at $p \leq .05$
Animal Learning	9	100%	89%
Human Abilities	10	40%	20%
Psychophysics	9	33%	33%
Reaction Time	3	67%	33%
Inkblot Tests	5	80%	80%
Laboratory Interviews	6	83%	33%
Person Perception	64	36%	23%

From this table, it is clear that the person perception task is least susceptible to the effects of experimenter expectancy. This is interesting because it is primarily these experiments that were objected to by Barber and Silver (1968) as being inappropriately analyzed. If Barber and Silver are correct, then the actual percentage of true expectancy effects obtained in these experiments should be considerably lower than the 23% listed above.

Expectancy Effects in Real-Life

As indicated, much of the laboratory research on expectancy effects has been subject to criticism. The recording of data, the analytical techniques, the ambiguity of some of the tasks - all of these have been offered as reasons to question the validity of published reports of expectancy effects. There is other evidence however - evidence from research on real-life situations.

Stuart Rice (1929) provided an early illustration of expectancy effects in his study of interviewer characteristics and interviewee responses. He observed a close similarity between interviewer expectancies and the nature of the responses obtained. Rice concluded that the expectancy of the interviewer was somehow communicated to the interviewee who responded accordingly. For example, interviewers who were prohibitionists obtained a greater number of responses from the interviewees ascribing their plight of dependency and hardship to "liquor" than did non-prohibitionists. While

there may be other explanations for the "how" of the results (e.g., errors of recording), Rice's observations nevertheless support the self-fulfilling prophecy concept.

Wyatt and Campbell (1950) also performed a study of interviewer expectancy effects. Before conducting their interviews (of 1948 Presidential Campaign), interviewers were asked to forecast the percentage response distribution they would obtain to each of five questions. While achieving statistical significance (P .02) in the case of only one question, for four of the five questions the interviewers tended to obtain more answers in the direction of their expectancy.

It has been established that the nature of care given to the mentally ill varies with the socio-economic class of the patient (Hollingshead and Redlich, 1958). Lower-class patients generally receive more institutional and custodial care than middle and upper-class patients. Similarly, Haase (1964) has demonstrated the effects of examiner social-class bias in the interpreting ofRorschach protocols; lower-class characteristics tend to receive less adjusted ratings. These kinds of data, while only suggestive, hint strongly at the possibility that a self-fulfilling prophecy is operating with regard to social-class in the prognosis, diagnosis and care of the mentally ill.

Studies of the placebo effect can also be considered as supporting the notion of a self-fulfilling prophecy. For example, Beecher (1966) compared the effects of morphine to the effects of placebo in the control of pain. He used a double-blind design in which neither the subject nor the investigator knew which drug had been administered. His results showed that morphine was no better at controlling pain than was the placebo. This was inconsistent with the work of other investigators who obtained data differentiating the placebo's effect from morphine's effect. Beecher concluded that his lack of knowledge of which drug was being administered made it impossible for him to communicate his expectation to the subject. The other investigators did not employ a double-blind design and therefore always knew which drug was being administered thus enabling the self-fulfilling prophecy to operate.

Another report by Beecher (1961) describes even more dramatic results. A surgical procedure was developed for the relief of angina pectoris. Although the results of this procedure were generally excellent, it was noted that surgeons who were enthusiastic about the procedure had considerably greater success in bringing relief to their patients than did those surgeons who were skeptical about the procedure. It was subsequently determined (experimentally) that the simple exercise of making skin incisions was just as effective in providing relief for angina as was the specific operation.

The patients, so it seems, were being affected by the surgeon's enthusiasm (or lack of it) for the operation.

In searching for commonalities among studies of interpersonal expectancy effect, it seems that the one characteristic that remains constant is the nature of the relationship between the expecter and the expectee. The self-fulfilling prophecy seems to operate in situations involving a command or superior - subordinate relationship. It may well be a requirement that he who has the expectation must have sufficient opportunity, credibility and authority to effect the behavior of the other. Consider the various interpersonal relations where the expectancy effect has thus far been demonstrated - experimenter/ subject, physician/patient, interviewer/intervieweee. All of these meet the requirements of a command relationship. If we apply this theory to other relationships, we should expect to find strong evidence for the self-fulfilling prophecy in the industrial (superior/subordinate), military (officer/ enlisted man) and classroom (teacher/student) worlds.

There are no reported studies of the expectancy effect in the military. And there is but one undocumented report of its effect in an industrial setting. Bavelas, in a personal communication to Rosenthal and Jacobson (1968A) describes a study where foremen were falsely led to believe that certain new employees scored high on tests of intelligence and finger dexterity and that certain others had scored low. After a

period of time, the foremen's performance appraisals and the production records for each of these employees were studied. The foremen evaluated more favorably the workers who were believed to be superior on the basis of the alleged test scores than those who were believed to be inferior. While this can be considered a result of a "halo" effect (criterion measure is not independent of the expectation) the additional finding that the objective production records also lined up with the expectations established adds credence to the likelihood that the self-fulfilling prophecy was operating.

While Bavelas' study is interesting, it lacks documentation and therefore cannot be considered as anything more than an anecdotal clue to the generalizability of the self-fulfilling prophecy. Additional research is needed to demonstrate the application of the concept to the industrial environment.

Unlike the military and industrial spheres, expectancy effects in the classroom have received considerable attention. Rosenthal's most popular work is a study conducted with Lenore F. Jacobson (1968A) on the effects of teachers' expectations upon pupils' intellectual competence. The operational intent of the study was to create in the classroom expectations that would be based solely on information presented to the teachers and not on judgments based on previous observations of the children involved. To accomplish this, the teachers of one predominately white

lower-class school in South San Francisco were told that further validation was required for a new test designed to predict intellectual gain in children. In May 1964, the teachers administered the test to all the children in kindergarten and grades 1 through 5. The following September, approximately 20 percent of the children (about 5 per classroom) were randomly designated as potential academic bloomers. The teachers were led to believe the designation came about as a result of the students' previous performance on the test. Thus the difference between the designated and the undesignated children was entirely in the minds of the teachers. All of the children were tested again four months after school started, at the end of that school year and May of the following year.

A discussion of the results of this study is influenced by growing criticisms of Rosenthal and Jacobson's procedures in analyzing and presenting the data (Snow, 1969; Thorndike, 1968). Besides pointing out the need for sound research in this area, these criticisms make it difficult to properly assess the findings. It is probably best to make the general statement that the study "suggests" that a self-fulfilling prophecy was successfully manipulated - - children from whom teachers expected intellectual gains seemed to exhibit such gains. For example, at the end of the school year, the designated children, especially of the younger grades, were shown as achieving a substantially greater gain in total IQ and in reasoning IQ (one of the two subtests) than did the

control group children. No appreciable difference was found in verbal IQ, the remaining subtest.

It should be noted that this basic experiment was subsequently repeated in two additional school systems in the Midwest and on the East coast with students from a different social class (middle class). These investigations (Rosenthal and Evans, 1968 and Conn, Edwards, Rosenthal and Crowne, 1968) did not show the same results. For example, in the Rosenthal and Jacobson investigation only the experimental girls showed significantly greater gain in reasoning IQ; in the Rosenthal and Evans study only the experimental boys showed significantly greater gains in reasoning IQ; and in the third study, both the experimental boys and experimental girls showed significantly greater gains in reasoning IQ than did their respective control group counterparts (but not in total IQ). These findings, while suggestive of the potentially powerful influence of teacher expectations, also indicate that the specific effects are complex and greatly affected by such factors as student's sex, social class and probably other variables as well.

Most of the experiments on teacher expectancy effects have used intelligence or scholastic achievement as the performance measure. Rosenthal (1969) describes Burnham's unpublished study (1968) of a successful demonstration of the hypothesis using a radically different criterion measure, namely, swimming ability. His subjects were boys and girls aged 7 through 14 who could not swim at the beginning of the two-

week experiment. The camp staff was told that half of the children showed unusual potential for learning to swim as determined by a battery of psychological tests. These children were, of course, selected as "high potential" on a random basis. At the end of two weeks the children were retested on the standard Red Cross Beginner Swimmer Test. As predicted, the children who allegedly had "high potential" showed greater improvement (P .005) than did children of the control group.

Not all of the investigations of teacher expectancy have been successful in demonstrating a statistically significant result. For example, Pitt (1956) found no effect on objective achievement scores of adding to or subtracting 10 IQ points from the student's actual IQ score. Both Claiborn (1968) and Rosenthal and Anderson (1969) were reported (Rosenthal, 1969) as finding negative results; i.e., a tendency for the control group children to gain more in total IQ than did the "potential bloomers" (P - .08). Of the thirteen teacher expectancy studies summarized by Rosenthal in his latest review (1969), only five had findings that were supportive at a probability level of .05 or below.

Communication of Expectation

Despite the criticisms and inconsistent results, there is enough evidence in the literature to suggest that the self-fulfilling prophecy may operate in interpersonal relations. The question now is how does it operate?

According to Rosenthal and Jacobson (1968B), the process by which one person communicates his expectancy to the other person is a subtle and unintended one. Laboratory studies have shown that visual cues such as facial expressions and posture and auditory cues (tone of voice) play a principal role in communicating one's expectancy to another. Adair and Epstein (1968), for instance, have demonstrated that the self-fulfilling prophe can come about in the person perception task through instructions previously tape-recorded by the experimenter. Despite the absence of visual cues (the experimenters were not present) expectancy effects were observed - - auditory cues must therefore have transmitted the experimenter's expectancy.

Despite the evidence presented by Rosenthal (1966, 1969) in favor of unintentional communication of the expectation, there is much to suggest that the behavior may be intentional. As indicated earlier, Barber and Silver (1968) have pointed out that much of the data obtained in the research on experimenter expectancy could have resulted from intentional cueing of subjects (e.g., prodding rats, smiling at correct responses, etc.) and intentional fabrication or mis-recording of data. Given the fact that there were several observations of cheating (e.g., Rosenthal and Fode, 1963A and Rosenthal and Lawson, 1964) and that the opportunity to intentionally cue or cheat was present in most of these studies, it is reasonable to question Rosenthal's assertion of solely unintentional communication. Further studies are needed to clarify which

of the modes of mediation (intentional or unintentional) plays the most significant role in producing the self-fulfilling prophecy.

When we move to classroom studies of expectancy effects the view that communications are subtle seems to be dubious. While there are no direct observational studies of teacher behavior, even Rosenthal & Jacobson (1968A) imply that the communication process is hardly subtle:

> "On the basis of other experiments on interpersonal self-fulfilling prophecies, we can only speculate as to how teachers brought about intellectual competence simply by expecting it. Teachers may have treated their children in a more pleasant, friendly and encouraging fashion when they expected greater intellectual gains of them.....Teachers probably watched their special children more closely, and their greater attentiveness may have led to more rapid reinforcement of correct responses with a consequent increase in pupil's learning" (pg. 180)

Again, experience tells us that a teacher communicates her expectations by way of such molar behavior as; the number and quality of interactions with the student, her general friendliness, her verbal expressions. It would be difficult to classify these as "subtle" behaviors.

Whether teachers fulfill their expectancy through behavior that is intentional or unintentional is another matter. Is the biased teacher unconscious of the fact that she is behaving in accord with her bias or is her behavior planned to satisfy that bias? It is an intriquing question; one that requires investigation.

One of the most interesting process variables put forth by Rosenthal and his students as having an impact on expectancy effects is that of "early data returns" (Rosenthal, Persinger, Vikan-Kline, and Fode, 1963 and Rosenthal, Kohn, Greenfield and Carota, 1965). Under this concept it is suggested that the early results of an expectancy experiment affects subsequently obtained data. When the early data returns confirm the expectation, subsequent data tends to also be confirming. If the early results are contrary to the expectancy, then the later data tends also to be contrary. Apparently, the early data serves as a "reward" or as a "punishment" for the prediction depending upon the nature of the returns. If they are confirming, the prediction is reinforced and interacting behavior with subsequent subjects tends to facilitate further confirmation of the expectancy. If, on the other hand, the early results are contrary to the prediction, then the prediction is weakened and subsequent subject interaction tends to hinder (or doesn't facilitate) the fulfillment of the prophecy.

Although the theory of early data returns seems quite plausible, there is some conflicting evidence. Barber and Silver (1968) for one, criticize the misleading statistical procedures of the few studies that supposedly show the effect. Rosenthal and Jacobson (1968) assert that "several studies showed that prophesied results became more likely as more subjects were contacted by each experimenter. In

fact, there was very little expectancy effect in evidence for just the very first seen subjects" (page 30). According to the notion of early data returns these initial failures should have resulted in subsequent failures to produce an expectancy effect.

Rosenthal (1969) explains these contrary results in an interpersonal learning context. He feels experimenters have to learn to cue the expected response. Thus with early subjects the experimenter is undergoing a "trial and error" phase which shapes his subsequent behavior with later subjects. In effect, the experimenter learns how to communicate his expectation. Under this theory, experimenters should be more successful in their influencing of subjects contacted later in the experiment. Rosenthal (1969) tells of six studies that provide evidence supporting this hypothesis.

The contradictory nature of these findings heightens our interest in early data returns. Are they really critical in determining whether or not subsequent results exhibit expectancy effects? If they are, it would have immediate implications for focusing upon early teacher-student interactions as the pivotal point in their relationship. By careful manipulation of early experiences it may be possible to maximize the positive effects of the self-fulfilling prophecy (i.e., where teacher has positive expectations) and minimize or eliminate the negative effects (i.e., where

teacher has negative expectations).

The Self-Fulfilling Prophecy and the Disadvantaged Student

Most of us will not be surprised to discover, that by every conceivable measure, children of low socio-economic status (the disadvantaged) do not perform as well in school as children of the middle and upper classes. The evidence is overwhelmingly supportive of the hypothesis that lower-class children are academically inferior to their counterparts of higher status (e.g., Coleman, 1940; Wilson, 1963; Doddy, 1963; Dreger and Miller, 1968; Sexton, 1961; Curry, 1962). There is no doubt that the poor, the "ghettoites", the disadvantaged, the culturally deprived, or whatever else you may want to call them are, indeed, inferior performers in school. But the question remains, what is it about being disadvantaged that lends itself to academic failure? What causes the lower-class child to perform worse than the middle class child?

Over the years there have been many theories and arguments as to the causes of academic failure amongst the lower class. There have been biological or hereditary explanations focusing on: inadequacy of brain size (Putnam, 1963); inherited deficiencies in intelligence (Tanser, 1939); and intellectual and neurological deficiencies resulting from premature births (Pasamanick and Knobloch, 1958).

Other explanations are concerned with the environment of the

poor. Sub-standard housing, poor nutrition, a lack of privacy, limited opportunities to explore the "outside" world; a scarcity of books, toys, and puzzles and a general pattern of parental apathy toward education are the typical factors referenced as causes of poor school performance. Taken together, these explanations focus upon a deprivation of the stimulus conditions required to function adequately in school.

Child-rearing explanations have centered on: the effects of an absentee male parent (Deutsch and Brown, 1964); the effects of the lower-class parental style of authoritarianism and physical discipline; and the effects of inappropriate (non-verbal) modes of expression (Miller and Swanson, 1960) and communication patterns (Bernstein, 1960; Milner, 1951).

The explanations offered, thus far, have focused upon causes which tend to be generated directly from the life of the child and his interaction with his parents. There is yet another class of explanations - those that relate to the school and especially, the teacher. Schools serving low-income students as compared to middle-income schools have been noted as deficient in regard to: physical facilities, amount of expenditures, quantity and quality of teacher, variability of curriculum and number and quality of remedial services offered (Sexton, 1961). As detrimental as they may be; high turnover, overcrowded classrooms, teaching inexperience,

unimaginative curricula and the like may be minor factors when considering the possibility that a self-fulfilling prophecy may be operating in these schools (Clark, 1965 and Deutsch, 1967).

It is quite likely that teacher expectations for the disadvantaged are that they cannot learn because of poor heredity, poor home background, "cultural deprivation", low IQ, lack of motivation and so on. When the children are tested, their low scores on inappropriate tests of aptitude and achievement reinforces this conviction; further reinforcement comes from the school system itself which tends to maintain a custodial frame of reference (as opposed to a teaching frame of reference) insofar as the education of lower-class or disadvantaged children are concerned. As a consequence of low expectations, the teachers (and society as a whole) demand less from the student in the way of academic achievement and often give up in their efforts to teach. The ultimate result is pupils who do poorly or fail - - just as the teacher expected.

Educators and researchers are becoming increasingly aware of the possibility that the self-fulfilling prophecy may be a major determinant of success or failure in school. Clark (1965), in his description of Harlem schools, puts it this way:

> "The most insidious consequence of these assumptions (assumptions that Negroes cannot learn) is that they are self-fulfilling prophecies. The fallacy in the assumptions does not mean that a system based upon them will be demonstrated to be ineffective; for once one organizes an educational system where children are placed in tracks or where certain judgments about their ability determine what is done for them or how much they are taught or not taught, the horror is that the results seem to justify the assumptions.....Children who are treated as if they are uneducable almost invariably become uneducable. This is educational atrophy." (page 128)

My purpose in closing the literature section of the paper with a discussion of the role of the self-fulfilling prophecy in the education of the disadvantaged is to emphasize the impact expectancy effects are believed to have. Experiments in the laboratory tend to sterilize the theory under investigation - the methods and the data become of central interest and not the implications. We are dealing here with a variable that may be affecting the destiny of thousands of students as they interact with their teachers. It is critical therefore that we learn as much as possible about its mechanisms and its applications.

Research Objectives

As we have seen, there is much criticism of the research dealing with interpersonal expectancy effects. The validity of this research is being severely questioned. In addition, most of the investigations of teacher expectancy effects have focused on white, middle-class children as the subject population. There are no reports of investigations utilizing adults as subjects

nor has there been any major effort to examine interpersonal expectancy effects in situations where: (1) the students are wholly from minority group populations and (2) the teaching occurs outside of the typical school environment.

Objective 1 - To test the validity of the self-fulfilling prophecy through an examination of its effects upon the learning performance of disadvantaged persons employed in industry.

The significance of teacher expectancy research is heightened by the ever-present educational assumption that classroom performance is predictive of out-of-school performance. Thus, if a student is a high achiever, or can be "made" into a high achiever via teacher expectations he is likely to do as well in other real-life achievement settings (e.g., at work). The studies conducted thus far have not tested this assumption in regard to achievement resulting directly from the self-fulfilling prophecy - - probably because of the longitudinal requirements of such research.

The industrial training environment presents a perfect opportunity to assess the longer-range effects of the self-fulfilling prophecy. Here the student learns his skills and immediately puts them to use. If a positive interpersonal expectancy effect is demonstrated in

training and the training is relevant to on-the-job behavior, then the recipient of the expectancy should be better trained and should thus perform better on the job. Since the teachers would not be interacting with the students on the job such a finding could not be construed as a direct effect of teacher expectations. Instead, it would support the view that teacher expectations, while having a direct effect upon training performance, may also have an indirect effect upon subsequent on-the-job performance. This conclusion would underscore the vital role teacher expectancy plays in the whole "education-job" cycle.

> Objective 2 - To assess the indirect influence of teacher expectancy through an examination of its effects upon related out-of-classroom (on-the-job) performance.

It was previously noted that the research on interpersonal expectancy effects presents conflicting theories as to the influence of "early data returns" upon subsequent effects of the expectation. While it appears logical to assume (from a learning theory standpoint) that early observations of the success of a prediction will lead to even greater successes with more experience, there is evidence to suggest (Rosenthal, 1968A) that, in demonstrations of the self-fulfilling prophecy, the initial data is often contrary to the expectancy.

Objective 3 - To assess the impact of early teacher experiences with expectancy effects upon subsequent expectancy situations.

RESEARCH ENVIRONMENT

The study described here focuses on a real-life (as compared to laboratory) environment. This means that the students, teachers, training, materials and almost everything else utilized in the study was functioning for reasons beyond the objectives of the research. As a result, E had to conduct the study with the understanding that there could be no interference with the purposes or procedures of the real-life system. It was not possible, therefore, to disrupt that system through added controls in any major way. This lack of experimenter control had its principal impact on E's attempts to maintain a consistent procedure for all subjects. As will be shown, E was not always successful in those efforts.

The system referred to above is a system for hiring and training the hard-core unemployed in the First National City Bank of New York. It is commonly referred to as the Canal Street Training Center (see Appendix A for a general description). Financed in part by the Department

of Labor, the program is designed to enable disadvantaged persons to fill typist and clerical positions within the Bank. The great majority of participants in the program (approximately 75%) are female and black. To be accepted into the program, applicants must meet the following requirements:

(1) certification by State Employment Service or Concentrated Employment Program as hard-core unemployed.

(2) passed the Bank's English Comprehension test at a 6th grade level.

(3) no police record.

Successful applicants are referred in groups of 20 (approximate) to the Canal Street Training Program which lasts, on the average, about 15 weeks. The curriculum consists of five subject areas: typing, English, mathematics, office machines and office procedures. There are approximately 15 teachers in the program. One teacher from each of the subject areas is assigned to the group for the duration of the program.

During the first day at the Training Center, several hours are spent in obtaining pre-training test data. To measure change, tests are administered a second time during the students' last two weeks at the Training Center.

In addition to these tests, performance in training scores are calculated for each of the five subject areas. The procedure used here is one where the teacher assigns performance points for various sub-topic scores which are then multiplied by their assigned weights. For example, in typing, lessons 1 - 30 has a weight of 3; if a student gets a test score of 90 - 100 on those lessons she gets 6 base points, which, when multiplied by 3, gives her a total of 18 points for that sub-topic.

Approximately two months after placement, follow-up evaluation forms appraising the on-the-job performance of the graduates are completed.

Hypotheses and Implications

Based on the objectives and considerations stated above, the specific hypotheses of this study can be stated as follows:

Hypothesis I - Trainees for whom a positive expectancy set is established will demonstrate significantly (.05 level of probability) better performance in training than their control counterparts (no expectancy set) as indicated by differences in:

a) performance on a battery of curriculum relevant achievement tests.

b) grade-point scores assigned by the instructors.

c) time spent in training

d) rate of turnover

e) attendance and punctuality

Hypothesis 2 - Trainees for whom a positive expectancy set is established will demonstrate significantly (.05 level) better performance on the job than their control counterparts as indicated by differences in:

a) supervisor ratings of actual job performance

b) attendance and punctuality

c) turnover rate

Should the results of the study support the hypotheses, it would have very real significance for the education of disadvantaged persons. Millions of dollars are being spent each year by federal, state and local government and private industry to educate people from ghetto areas. Almost the entire focus is on the curriculum, materials, tests and services with little attention paid to the teacher. In view of what is already known about the self-fulfilling prophecy, perhaps a re-assessment of priorities is in order. With proper application of our knowledge, perhaps, through learning and selection, we can eliminate negative expectancies; develop positive expectancies and, in so doing, improve the success probability of programs for the disadvantaged in industry, schools or anywhere.

Specific implications for further research would include;

(1) A need to investigate the ways by which teachers are able to effect changes in performance which satisfy their expectancies.

(2) A need to study the qualities and characteristics of teachers that facilitate actualization of a self-fulfilling prophecy.

(3) A need to examine the qualities and characteristics of students (subjects) that enhance susceptibility to a self-fulfilling prophecy.

(4) A need to teach teachers about their "power of prophecy".

(5) A need to assess and control for the impact of expectancy effects in programs (educational, health, psychiatric, etc.) involving interaction with minority or disadvantaged group members.

METHOD

Subjects

Six consecutive classes of new employees attending the Canal Street Training Program during June - August 1969 comprised the entire study population - 130 trainees in all. The experimental group, the group for which a positive expectancy set was established, consisted of 30 randomly selected trainees and the remaining 100 were the control (no expectancy set). Both groups were predominately female - 93 and 91 percent respectively. In order to minimize teacher suspicions that an experiment was being conducted, the number of students selected and reported to the teaching staff as "having outstanding potential" (positive expectancy set) was varied from class to class. The first class had five such trainees, the second class four, the third class six, and so on.

Independent and Dependent Variables

The manipulation of a positive expectancy set constituted the independent variable of this study. Operationally, this refers to written and oral communications to the teaching staff attesting to the alleged "outstanding potential" of several students.

The dependent variables were performance measures which were simultaneously being used to assess the effects of the Canal Street Training Program. These were:

1. Total grade-point scores in each of the five subject areas.

(See Appendix B for form used in data collection.)

2. Gain or post-test scores on various standardized achievement tests administered pre and post training. (See Appendix C for test battery.) It should be noted that each of these tests are related to the curriculum of the training program and are therefore relevant indices of the impact of teacher expectancy effects.
3. Number of days spent in training, number of days absent, number of times late and whether or not trainee drops out or is terminated from the program.
4. Number of days absent, number of times late and whether or not trainee is voluntarily or involuntarily terminated during first two months on the job. (See Appendix D for form used in data collection.
5. Supervisory ratings of job performance after approximately 60 - 70 working days on the job. (See Appendix D for form used in data collection.)

Procedure

The procedure described below was replicated for each of the six classes of trainees except as indicated.

Testing - In accordance with the usual

practice at the Canal Street Training Center (CSTC), the students were tested during the first week and last two weeks of the training program.

Establishing the Expectancy Set. At the outset of the experiment, top-ranking Bank officials of Training and Education (accomplices of E) held a meeting with the administrative staff (i.e., Director and Assistant Director) of the CSTC. The administrators were falsely informed that the Fundamental Achievement Series (one of the tests given at the CSTC) was shown, in the literature, to have been successful in predicting outstanding performance during training. An abstract of an alleged research study (Appendix E) was distributed in support of the statement.

E pretended interest in developing a new curriculum for the better students.* To help accomplish this he suggested a classroom behavior study of a sample of high potential students as identified by the Fundamental Achievement Series. After some discussion, it was agreed that the instructors of the next six starting classes were to be told the names of the "high potential"

*E, as the Bank's Director of Training Evaluation, was responsible for conducting such research and was well known to the Center staff.

students of those classes and were to observe them closely for characteristics that would help in the identification of an appropriate training program (e.g., attitudes, behavior, personality). It was arranged that at some point during the training period E, or a member of E's staff would return and interview the instructors on their classroom observations.

Two days after pre-testing, E conducted individual meetings with the instructors.* They were informed of the purpose and procedures of the alleged study and were told about the supposed predictive value of the Fundamental Achievement Series. Finally, they were given the names of the "outstanding potential" trainees of their class (Appendix F).

It should be noted that there were considerable differences among the teachers in their initial attitude towards the alleged research. Some were quite enthusiastic, others were relatively neutral and a few were suspicious of or hostile towards the study. Three of the thirteen teachers involved mentioned to E that the research might produce a self-fulfilling prophecy and wondered out loud if that was the true intent of the study. After

*There were 13 teachers involved in the study. Ten were white (9 females, 1 male) and three were blacks (2 females, 1 male). One of the instructors was of Puerto Rican descent.

considerable discussion, all three of them appeared to be convinced that E was "honest" in his study and that there were no secret objectives. One other instructor was very hostile but this seemed to be more a characteristic of the teacher's personality than an objection towards the project. For example, she was the only instructor who spoke negatively about the caliber of the students in the program. All of the others expressed a generally positive attitude towards their work, the program and their students.

Many of the teachers were assigned to more than one class. In these instances the names of the later class "outstanding" students were mailed to the instructors unless E was uncertain of a teacher's cooperation. Whenever there was doubt E held a second interview (but never a third interview) with the instructor in an attempt to reinforce the expectancy set and to stress the importance of the alleged study. An additional action intended as a reinforcement of the expectancy occurred 3 weeks after the last class had been pre-tested. A letter was sent (Appendix G) to each instructor specifying, as a summary, the names of the "outstanding potential" students of all six classes.

Verifying the Expectancy Set. Eight weeks after the last class had been pre-tested, an associate of E (an employee of the bank working for E) conducted individual interviews with each of the thirteen instructors participating

in the study. The instructors were led to believe that the interviews were for the purpose of obtaining their observations and generalizations concerning the classroom characteristics of the "outstanding potential" students. The true objective, however, was to determine the extent to which the expectancy set manipulation was successful.

The interviews were conducted at the Canal Street Center; each consisted of a 15-30 minute period of open-ended questioning of the instructor's observations and experiences. Appendix H shows the instrument used by the interviewer in guiding the discussion.

Since the students were still in training it was necessary to guard against exposure of the true purpose of the research. The interviewer was therefore instructed <u>not</u> to probe the teachers directly for their attitudes toward the study. It was felt that such attitudes would be uncovered, without solicitation, in the course of their commentary about their experiences with the "outstanding potential" students. The likelihood that this would occur was enhanced by E not participating in the interviews. E's position of authority within the bank was seen as being a possible inhibitor of the instructors' willingness to "speak their mind". E's confederate, on the other hand, was an employee of almost equal status to the instructors and, as such, would be more likely to generate an interview environment conducive to honest and open discussion.

Three of the thirteen teachers interviewed* were suspicious of the true intent of the study (two of the three were equally suspicious when E first told them about the project). All three seemed to think that a self-fulfilling prophecy was being manipulated and as a result were quite negative towards the research. The interviewer kept his role by neither confirming nor denying their suspicions. The other ten instructors seemed to have a positive attitude towards the study as indicated by their acceptance of the "outstanding potential" concept and willingness to discuss personality and behavior characteristics of the "outstanding" group. Most of them, however, indicated that the FAS test was not an infallible predictive instrument. All but four of the instructors expressed the view that "others should have been on the list."

*Two were black females and one was a white male.

RESULTS

The findings of the study are presented below under two major headings; (1) the effects of inter-personal expectancy in the learning (training) environment and (2) the effects of inter-personal expectancy in the working (on-the-job) environment.

(1) Inter-personal Expectancy Effects during Training

Six dependent measures were employed in this phase of the study - scores on a standardized achievement test battery, classroom grade point scores, time spent in training, rates of turnover, absenteeism, and punctuality.

(A) Standardized Achievement Test Scores

A complete test record for each of the seven achievement tests used in the study consisted of a pre-test raw score, a post-test raw score and a computed gain score (i.e., post test score minus pre-test score). As indicated on the tables that follow, the number of complete test records actually obtained varied from test to test and in all instances was well below the number of trainees (130) in the study population. These discrepancies are primarily a result of the substantial number of students (32) who were not available for post-testing. Twenty-nine of these were program dropouts who either requested or were asked to terminate from the program (i.e., the bank). The remaining three students were absent at the time of testing.

Since testing was an integral part of the Canal Street Training Operation, it was not possible for E (without uncovering the true intent of his study) to exercise any control over the testing procedures. As a result, on several testing occasions (pre and post), it was arbitrarily decided by the testing personnel to omit certain tests. This usually happened when there were time constraints during testing or when the supply of test materials was depleted. These instances explain the different "N's" among the seven tests.

Analysis of Overall Effect of Inter-Personal Expectancy. Table I presents the results of t-test comparisons of the gain scores for each of the seven tests. As shown, there are no significant differences between the total control and experimental groups on any of the tests. If anything, the trend of the data favors the control group (in 6 of 7 tests, the control group mean was higher). To insure that these findings were not affected by differences in baseline data, a comparison of pre-test scores (for trainees with complete test records) was computed for each test. Table 2 shows that there were no significant differences in pre-test performance. Thus the gain score findings appear to be an accurate indication of

TABLE I

RESULTS OF *t* TEST COMPARISONS OF GAIN-SCORES ON THE ACHIEVEMENT TEST BATTERY

TESTS	GROUP	N	MEAN	S.D.	*t**
FAS-VERBAL	E	22	6.05	3.62	.53
	C	44	9.25	7.40	
FAS-NUMERICAL	E	22	8.33	7.00	.15
	C	42	9.57	7.51	
DAT-NUMERICAL	E	6	1.67	5.06	1.54
	C	17	4.23	3.15	
NUMBERS	E	23	7.83	5.26	.09
	C	52	7.15	5.66	
OFFICE TERMS	E	18	2.50	4.57	.51
	C	39	3.77	4.02	
ENGLISH COMP.	E	10	-.70	2.79	1.24
	C	30	.17	5.01	
VOCABULARY	E	11	2.09	3.18	.72
	C	30	3.60	4.34	

*All *t* values are not significant at .05 level

TABLE 2

RESULTS OF *t* TEST COMPARISONS OF PRE-TEST SCORES ON THE ACHIEVEMENT TEST BATTERY

TESTS	GROUP	N	MEAN	S.D.	*t**
FAS-VERBAL	E	22	79.27	5.87	.04
	C	44	76.05	9.14	
FAS-NUMERICAL	E	22	44.32	7.85	.00
	C	42	44.17	7.08	
DAT-NUMERICAL	E	6	6.17	5.01	.04
	C	17	6.47	4.48	
NUMBERS	E	23	16.87	5.67	.09
	C	52	18.59	7.74	
OFFICE TERMS	E	18	9.61	4.11	.07
	C	39	9.00	4.06	
ENGLISH COMP.	E	10	31.20	4.98	.03
	C	30	32.07	4.74	
VOCABULARY	E	11	10.91	1.98	.17
	C	30	9.37	2.42	

*All *t* values are not significant at .05 level

achievement gain.

Because of the small "N" in several of the tests and since the groups were selected on a random basis, it was decided to compare all post-test scores, without regard to whether or not a complete test record was available. This meant that the post-test data of students who had not been pre-tested would also be included in the analysis. The results of this comparison (Table 3) were consistent with the gain-score findings - - the differences between the two groups were not significant.

Analysis of Effect of Early Data Returns. The conflicting evidence regarding "early data returns" (see page 20 - 22) led to an examination of its impact in this experiment notwithstanding the lack of a main effect. It was conceivable, in view of the early data returns literature, that there were performance differences between the first seen and later seen subjects. If such differences did occur it might indeed provide an explanation for the non-significant results described above. In other words, the poor performance of the first seen subjects may have masked, in an overall test, the expectancy effects occurring among later seen subjects.

TABLE 3

RESULTS OF *t* TEST COMPARISONS OF POST-TEST SCORES ON THE ACHIEVEMENT TEST BATTERY

TESTS	GROUP	N	MEAN	S.D.	*t**
FAS-VERBAL	E	22	85.05	5.24	.01
	C	45	85.53	5.39	
FAS-NUMERICAL	E	22	52.32	8.11	.03
	C	43	53.93	6.62	
DAT-NUMERICAL	E	22	9.23	5.34	.10
	C	50	10.24	5.20	
NUMBERS	E	23	24.70	6.99	.05
	C	53	26.02	8.32	
OFFICE TERMS	E	18	12.11	4.81	.06
	C	40	12.85	5.31	
ENGLISH COMP.	E	22	30.86	5.55	.00
	C	51	30.78	5.61	
VOCABULARY	E	23	13.39	4.68	.00
	C	51	13.37	5.73	

*All *t* values are not significant at .05 level

In order to examine the influence of early data returns upon test performance it was first necessary to sub-categorize the study population into "first seen" and "later seen" groups. For the four tests associated with the English curriculum (FAS-Verbal, English Comprehension, Vocabulary and Office Terms) this was accomplished by keying on the English instructors' experience as follows:

class 1	-English Teacher X	= First Seen
class 2	-English Teacher Y	= First Seen
class 3	-English Teacher Y	= Later Seen
class 4	-English Teacher X	= Later Seen
class 5	-English Teacher Y	= Later Seen
class 6	-English Teacher X	= Later Seen

Thus, the students of the first two classes comprised the "First Seen" category and the students of the last four classes the "Later Seen" category. It should be mentioned that both of the English instructors involved in the study suspected the motives of E and voiced their concern about a self-fulfilling prophecy (see page 39). This, of course, has significance for the interpretation of results relating to English performance.

The study population was similarly sub-categorized for the three tests associated with the mathematics curriculum (FAS-Numerical, Numbers, DAT-Numerical).

class 1	- Math. Teacher A	= First Seen
class 2	- Math. Teacher B	= First Seen
class 3	- Math. Teacher C	= First Seen
class 4	= Math. Teacher A	= Later Seen
class 5	- Math. Teacher C	= Later Seen
class 6	- Math. Teacher A	= Later Seen

The students of the first three classes comprised the "First Seen" category and the students of the last three classes constituted the "Later Seen" category. As in the case of the two English instructors, Math. Teacher B expressed serious doubts about the self-fulfilling nature of the study.

Tables 4 - 7 present the findings of the t-test comparisons of the performance of first seen and later seen students. If the performance of the first seen students had a negative impact on the overall results, the means of the later seen experimental students should be significantly larger than the means of the first seen group. With the exception of the Office Terms test (Table 4), the comparisons do not show statistically significant differences.

The Office Terms result, at first glance, would seem to support the expectancy hypothesis. The experimental group mean (4.18) for later seen students was significantly larger

TABLE 4

COMPARATIVE PERFORMANCE OF FIRST SEEN AND LATER SEEN EXPERIMENTAL STUDENTS ON FOUR ENGLISH ACHIEVEMENT TESTS

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
TESTS	N	MEAN	S.D.	N	MEAN	S.D.	t
FAS-VERBAL (Gain-Scores	6	4.17	2.48	16	6.38	3.97	.53
ENGLISH COMPREHENSION* (Post-Scores)	6	30.83	6.09	16	30.88	5.32	.00
VOCABULARY* (Post-Scores)	7	14.14	5.72	16	13.06	4.10	.08
OFFICE TERMS (Gain-Scores)	7	-.14	4.32	11	4.18	3.88	30.28**

*Because of small Gain-Score "N's" in English Comprehension and Vocabulary tests, post-test scores were used.

**Significant at .01 level, all other values are not significant at .05 level.

TABLE 5

COMPARATIVE PERFORMANCE OF FIRST SEEN AND LATER SEEN CONTROL STUDENTS ON FOUR ENGLISH ACHIEVEMENT TESTS

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
TESTS	N	MEAN	S.D.	N	MEAN	S.D.	t**
FAS-VERBAL (Gain-Scores)	12	7.58	4.13	32	9.88	8.22	.30
ENGLISH COMPREHENSION* (Post-Scores)	18	31.56	6.12	33	30.36	5.26	.04
VOCABULARY* (Post-Scores)	18	11.00	4.93	33	14.67	5.71	.33
OFFICE TERMS (Gain-Scores)	19	2.68	3.03	20	4.80	4.54	.79

*Because of small Gain-Score "N's" in English Comprehension and Vocabulary tests, post-tests scores were used.

**All t values are not significant at .05 level.

TABLE 6

COMPARATIVE PERFORMANCE OF FIRST SEEN AND LATER SEEN EXPERIMENTAL STUDENTS ON THREE MATHEMATICS ACHIEVEMENT TESTS

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
TESTS	N	MEAN	S.D.	N	MEAN	S.D.	t**
FAS-NUMERICAL (Gain-Scores)	10	6.70	6.88	12	9.00	7.02	.34 (.10)
NUMBERS (Gain-Scores)	11	6.54	5.53	12	9.00	4.71	.38 (.07)
DAT-NUMERICAL* (Post-Scores)	10	9.50	5.70	12	9.00	5.02	.05 (.20)

*Because of a small Gain Score "N" in the DAT-Numerical test, post-test scores were used.

**Parentheses indicate the t-scores obtained when the supposed "biased" class was excluded from the analysis (see page 52). All t values are not significant at .05 level.

TABLE 7

COMPARATIVE PERFORMANCE OF FIRST SEEN AND LATER SEEN CONTROL STUDENTS ON THREE MATHEMATICS ACHIEVEMENT TESTS.

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
TESTS	N	MEAN	S.D.	N	MEAN	S.D.	t**
FAS-NUMERICAL (Gain-Scores)	14	5.64	8.08	28	11.54	6.36	1.04 (.53)
NUMBERS (Gain-Scores)	24	7.12	5.81	28	7.18	5.53	.01 (.11)
DAT-NUMERICAL* (Post-Scores)	21	10.10	5.48	29	10.34	4.98	.02 (.02)

*Because of a small Gain-Score "N" in the DAT-Numerical test, post-test scores were used.

** Parentheses indicate the t-scores obtained when the supposed "biased" class was excluded from the analysis (see page 52). All t values are not significant at .05 level.

(.01 level) than for first seen students (-.14) and the control group counterparts showed no difference. A closer examination of the data shows, however, that the mean gain of -.14 was largely a result of one extreme gain score (-8). Since the number of items in the cell was quite small (7) and as there were no significant differences for any of the other achievement in English tests, it is most probable that the Office Terms finding was a chance occurrence.

Because of the small number of entries in the experimental group cells, any findings in Tables 4 and 6 of significant trend data would tend to support the hypothesis. The Tables show that for five of the seven tests the means of the later seen experimental group were larger than the means of the first seen experimental group. The somewhat suggestive nature of this evidence is eliminated however by the finding that the control counterparts also show a five out of seven ratio (Tables 5 and 7).

It should again be mentioned that both English instructors and one Mathematics instructor made statements to E which indicated a suspicious attitude towards the study. The results obtained may very well be an indication that these three teachers never believed or accepted the expectancy set manipulation. To examine this possibility, the mathematics test data was re-analyzed without the data of the "biased" class. The results, however, still did not show any indica-

tion that the expectancy set was operating. (see parentheses under t column in Tables 6 and 7).

Summary. There is nothing in the data to suggest that the positive expectancy manipulation has had an influence upon student performance on a training related achievement test battery. In general, the "outstanding potential" students test results were no better or no worse than their control counterparts.

(B) Classroom Grade-Point Scores

The program of the Canal Street Training Center consists of two basic education courses: English and mathematics; and three industrial training courses: typing, office procedures and office machines. As in most educational systems, student achievement is determined by test results (standard tests given in the classroom at various times) and classroom performance. Using these criteria, the instructors calculate grade-point scores for the various sub-topics that comprise the courses (see Appendix B and page 30 for a description of the procedure). With the exception of Office Machines, the maximum score possible for the basic curriculum of each course is 90 points (excluding bonus points); 108 points is possible for Office Machines.

The statistical analysis of the obtained grade-point scores excluded several groups of data as follows:

(1) Partial grade-point scores for those students who

were terminated prior to the completion of the program (24 control, 5 experimental)

(2) Partial grade-point scores for those students who, because of management requirements, were placed on the job prior to "completion of the program" (14 control, 7 experimental). Based upon discussions with the program's administrative staff, it was decided that for the purpose of the analysis, seventy days or more in the Canal Street Center would satisfy the "completion of the program" requirement.

(3) Grade-point scores for one class (11 control, 4 experimental) because of identified irregularities in the computation of the scores. It should also be noted that this was the class taught by the "biased" mathematics instructor mentioned previously.

As indicated earlier, both English instructors were probably unaffected (did not accept) by the expectancy manipulation. The English grade-point data is therefore included in the analysis of an overall expectancy effect for comparative purposes only.

Analysis of Overall Effect of Inter-Personal Expectancy. As shown in Table 8, the average grade-point scores achieved by the experimental students did not differ significantly from the average scores of the control group for any of the five courses. The data provides no suggestion that the expectancy set had an impact upon student classroom performance.

TABLE 8

RESULTS OF t TEST COMPARISONS OF GRADE POINT SCORES
FOR STUDENTS IN TRAINING 70 DAYS OR MORE (N=14 Exp., 50 Cont.)

COURSE	GROUP	MEAN	S.D.	t*
ENGLISH	E	73.42	6.90	.00
	C	73.62	7.02	
MATHEMATICS	E	75.21	8.68	.01
	C	74.29	8.13	
OFFICE MACHINES	E	47.43	9.31	.01
	C	48.08	12.35	
OFFICE PROCEDURES	E	54.00	10.87	.11
	C	48.14	10.53	
TYPING	E	59.21	7.95	.19
	C	47.94	12.84	

*All t values are not significant

Analysis of Effect of Early Data Returns. An analytical procedure similar to that employed for the achievement test battery was used to examine the effect of early data returns upon grade-point scores. The class composition of the "first seen" and "later seen" groups were as follows:

Mathematics	1st Seen Group - classes 1 and 3 Later Seen Group - classes 4, 5 and 6
Office Machines	1st Seen Group - classes 1 and 4 Later Seen Group - classes 3, 5 and 6
Office Procedures	1st Seen Group - classes 1 and 3 Later Seen Group - classes 4, 5 and 6
Typing	1st Seen Group -classes 1 and 4 Later Seen Group -classes 3, 5 and 6

(Note: Because of insufficient cell entries, an analysis of the English course was not possible.)

Tables 9 and 10 present the results of an examination of the differences between the means of the first seen and later seen groups (experimental and control). Again, the findings are not significant nor is there any evidence of a trend in the desired direction. Although it is not my intention, at this point, to offer alternative explanations for these results, one obvious consideration is the very small size of the experimental groups.

Summary. Training achievement, as measured by grade-point scores, was shown to be uninfluenced by the positive expectancy set. Assuming the reliability and validity of the grade point scoring system, the experimental and control students were not distinguishable in the classroom.

TABLE 9

GRADE-POINT SCORE COMPARISONS BETWEEN FIRST SEEN AND LATER SEEN EXPERIMENTAL STUDENTS IN TRAINING 70 DAYS OR MORE.

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
COURSE	N	MEAN	S.D.	N	MEAN	S.D.	t*
MATHEMATICS	4	79.00	7.68	10	73.70	8.60	.07
OFFICE MACHINES	6	51.33	12.05	8	44.50	4.77	.13
OFFICE PROCEDURES	4	41.50	4.39	10	59.00	8.38	.42
TYPING	10	56.83	11.10	8	61.00	3.28	.07

*All t values are not significant at .05 level.

TABLE 10

GRADE POINT SCORE COMPARISONS BETWEEN FIRST SEEN AND LATER SEEN CONTROL STUDENTS IN TRAINING 70 DAYS OR MORE.

	FIRST SEEN STUDENTS			LATER SEEN STUDENTS			
COURSE	N	MEAN	S.D.	N	MEAN	S.D.	t*
MATHEMATICS	20	76.15	9.26	30	73.20	7.15	.04
OFFICE MACHINES	25	52.52	15.49	25	43.64	5.06	.17
OFFICE PROCEDURES	20	47.80	10.42	30	48.37	10.59	.01
TYPING	25	52.40	9.30	25	43.48	14.26	.17

*All t-values are not significant at .05 level.

(C) Time in Training, Turnover, Absenteeism and Lateness.

Up to this point, we have been unable to confirm the principal hypothesis of the study. Insofar as educational achievement is concerned this experiment does not demonstrate the predicted impact of inter-personal expectancy in the industrial classroom. We turn now to other educational achievement indices - those not uncovered directly in the classroom but found, instead, in the outcomes of the total training experience.

It is a basic responsibility of the entire Canal Street Training staff to informally orient the student to the world of work. Through day-to-day interactions with the instructional, counseling and administrative staffs, it is expected that the student will learn how to conduct himself in the business environment. The anticipated pay-off to the employer is:

(1) more rapidly qualified and productive trainees.
(2) lower turnover rates during training on the job.
(3) less absenteeism and lateness in training and on the job.

Under the expectancy hypothesis, we would anticipate that the interacting experiences between the experimental student and the staff (most staff members were knowledgeable of the "outstanding potential" students) would differ qualitatively and quantitatively from similar staff interactions with control group students. The label "outstanding potential"

should result in spin-off benefits to the student in the way he is treated and communicated with outside of the formal classroom. As a result, we would expect differential outcomes between the experimental and control groups in regard to the time they spend in training, the number that leave the program and the frequencies of absenteeism and lateness.

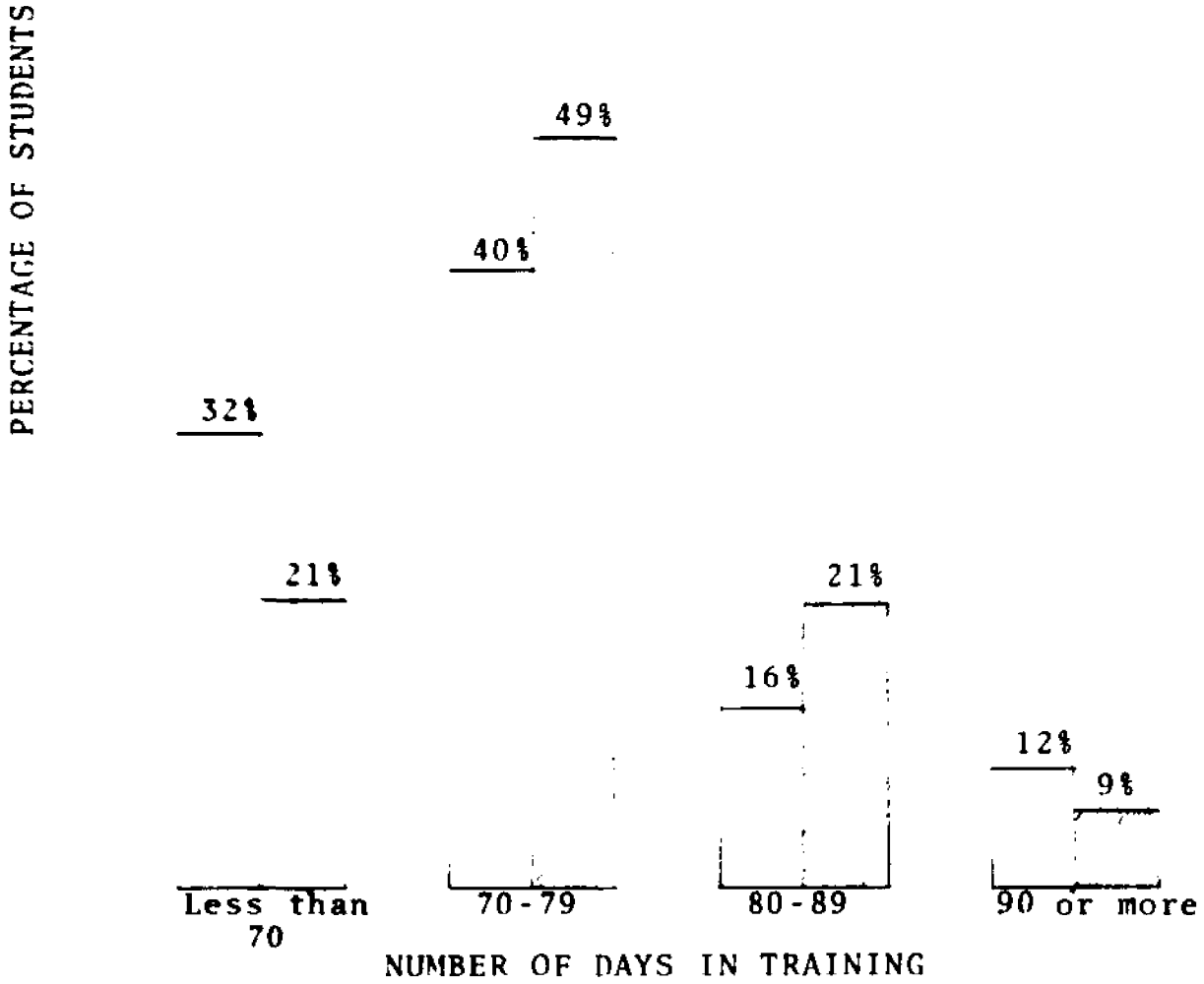
Time Spent in Training. The decision as to when a trainee should be placed on the job is made by the instructional and administrative staff at the center. The principal criteria employed in making this decision are: urgency to fill job openings, general classroom performance (not standardized test performance), out-of-classroom behavior and attitudes, instructor and counselor recommendations, length of time since entering the program and what can be called a general "gut" feeling about the student. Although there is no obvious pattern to the way these criteria are applied, it can be said that, in general, the student who gets placed on the job earlier is viewed by the staff as being the more "successful" trainee.

Figure 1 illustrates the percentage distribution of days spent in training by the experimental and control groups. As indicated, there is no difference between the two distributions insofar as the overall means (or medians) are concerned. The average number of days spent in training by both groups of students was approximately seventy-six.

FIGURE I

PERCENTAGE DISTRIBUTION OF DAYS SPENT IN TRAINING

	N	MEAN	MEDIAN
Exp.	25	75.6	76.0
Cont.	76	75.4	76.0



As indicated previously, 70 days in training was considered to be a minimum requirement to be numbered among those completing the entire program. Those students who were released from training and put to work in less than 70 days were, in the opinion of the staff, the "best" or most "likely to succeed". A total of 24 students were placed on the job in less than 70 days - 8 experimental, 16 control. Figure 1 indicates that these occurrences accounted for 32 percent (8 of 25) of the experimental group distribution and 21 percent (16 of 76) of the control group distribution. A test for the significance of the difference between these two proportions indicated a probability value of .26. Thus, the proportion of experimental students to leave the program early was considerably larger than among the control counterparts. The probability level of the differences at the other points of the histogram (Figure 1) were .45 higher.

Although not significant at the desired level (.05), these results provide the first indication that the established expectancy may have had an impact. The question remains however: Was the impact of a self-fulfilling nature or was it a "halo" effect? Under an interpretation of the self-fulfilling prophecy, it would be argued that, in general, the "outstanding potential" students, through their interactions with the instructors, demonstrated better classroom performance and better employee behaviors and attitudes than the control group students. Observation of these differences by the Training Center staff led them

to select proportionately more experimental students than control students for early work assignments. Such an interpretation is weakened by the fact that all of our classroom achievement data (grade point scores and test scores) points to a conclusion that there were no performance differences between the two groups. Furthermore, there were no classroom achievement differences between the 24 trainees who left the program in less than 70 days and the rest of their respective groups.

It is also possible that the proportionate difference in the number of early assignments was a direct response by the decision-makers (administrators, instructors) to the knowledge that there were "outstanding potential" students in the group (i.e., halo effect). When confronted with an "outstanding" student the decision-making process could have gone something like this:

> "I know she has outstanding potential. I also know that she did not excel in class. She probably did not try very hard or maybe the instruction was inadequate. She will do very well on the job, I'm sure. Let's get her out of training and on to a job."

While such halo effect decisions tend to confirm the view that the expectancy set manipulation was successful, it disputes the notion that the results of the expectancy were self-fulfilling in nature. As long as the criterion measure is determined by inputs of the individuals who were given the expectancy, the findings can always be interpreted as

a conscious or even sub-conscious halo effect. To be proven a self-fulfilling result, the criterion, as in Rosenthal's work, must be independent of the recipient of the expectancy input. At the Training Center the decision-makers (instructors and administrators) were aware of who the "outstanding" students were. It is quite possible that the proportionate difference of early assignments favoring the experimental group was a direct result of that awareness and had no basis in fact.

Turnover Rates. The overall turnover rate during training for all students participating in the study was 22.3 percent (29 of 130). This compares favorably with the long standing Training Center turnover rate of 24.8 percent.

Table 11 presents a comparison of the turnover figures by treatment group (experimental/control) and by turnover type (voluntary/involuntary). As shown, the experimental group had a total turnover rate of 16.6 percent (5 of 30) as compared to the control group's 24 percent (24 of 100). The difference between these proportions is not statistically significant, but the positive direction of this difference is certainly suggestive.

An examination of the total turnover results by type of turnover reveals somewhat stronger support for the expectancy hypothesis. Table 11 shows that the proportionate number of students who involuntarily left the program (i.e., were fired) was considerably higher (.24 level of probability) for the control group (10 percent) than for the experimental group

TABLE 11

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR TURNOVER RATES DURING TRAINING.

	EXPERIMENTAL GROUP		CONTROL GROUP			
	N	% OF TOTAL N (30)	N	% OF TOTAL N (100)	z	p
TOTAL TURNOVER	5	16.6	24	24.0	.85	.39
INVOLUNTARY TURNOVER	1	3.3	10	10.0	1.16	.24
VOLUNTARY TURNOVER	4	13.3	14	14.0	.09	.93

(3.3 percent).

The observed differences in involuntary terminations could well be explained by the self-fulfilling prophecy concept. Under that theory, the "outstanding potential" input would have led to "better" communications between the staff and the student which, in turn, would have led to more desirable student behaviors and performance and finally, to a smaller percentage of fired students. The pattern suggested by the self-fulfilling theory takes on added significance when the reasons for the involuntary termination are explored. The one experimental student who was fired was in the program only for 4 days and, in fact, never really reported for training. Hence there was no real opportunity for the instructors to interact with her. Of the ten fired control group students eight were let go because of poor attendance or poor general performance and only two were fired because they failed to report. If we disregard the "failure to report" students, then the percentage of involuntary terminations becomes zero percent for the experimental group and eight percent for the control. That difference is at a .10 level of probability. It should again be pointed out however, that a halo effect interpretation of these results is possible. The staff decision to fire a trainee could have been influenced by the knowledge that certain students were "outstanding" and others were not.

Voluntary terminations (those that quit), on the other hand,

were beyond the control of the training staff. In general, trainees quit the Bank for reasons independent of their experiences at the Center. Some were moving out of the city, others returned to school, a few were having babies, and several had to stay home to care for their children. Since the motives for leaving were not associated with the student-teacher interaction, the impact of the expectancy set should not show up in the voluntary termination figures. As suggested, there was no difference in the proportion of voluntary terminations - 13.3 percent of the experimental group and 14 percent of the control group voluntarily resigned from the program.

Absenteeism and Lateness Rates. Table 12 shows that in regard to the overall absence and lateness rates there is no statistically significant difference between the experimental and control groups. There is a trend in the data, however, indicating that the experimental students were absent considerably less often than the controls (.17 level of probability).

A closer examination of the absence data (Table 13) points out a salient result - that there was a significant difference (.03 level of probability) favoring the experimental group in the number of unexcused absences (i.e., where the trainee is not paid for the day). As indicated, 23 percent of the control group absences were not excused, whereas only 12 percent of the experimental group absences were unexcused. Further, an examination of the proportion of unexcused absences to total days

TABLE 12

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR ABSENCE AND LATENESS RATES DURING TRAINING

	EXPERIMENTAL	CONTROL	z	p
TOTAL NUMBER OF DAYS ABSENT / TOTAL NUMBER DAYS POSSIBLE TO BE ABSENT	.0391	.0453	1.39	.17
TOTAL NUMBER OF DAYS LATE / TOTAL NUMBER OF DAYS POSSIBLE TO BE LATE	.0407	.0373	.76	.45

TABLE 13

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR UNEXCUSED ABSENCE RATES

	EXP	CONT	z	p
TOTAL NUMBER OF UNEXCUSED ABSENCES / TOTAL NUMBER OF ABSENCES	.1153	.2290	2.23	.03
TOTAL NUMBER OF UNEXCUSED ABSENCES / TOTAL NUMBER OF DAYS POSSIBLE TO BE ABSENT	.0045	.0105	2.65	.01

available to be absent shows the experimental group to have significantly fewer unexcused absences (.01 level). The decision to excuse an absence is made by the administrators of the Training Center. If a trainee calls in or otherwise reports his absence he will in most instances be paid for the day. Why then did the control group trainees report their absence less often than the experimental group students? One possibility is that they were not as well informed as to the requirement of absence notification. This would be consistent with a self-fulfilling prophecy interpretation. Another possibility is that the administrators were more lenient in applying the notification criterion to the "outstanding" group. This would be consistent with a halo effect interpretation. Under both explanations, however, the expectancy set is shown as having considerable influence on the results.

SUMMARY. Unlike learning achievement, there is some evidence in the measurement of time spent in training, turnover and absenteeism which supports the hypothesis of the study. Those students for whom a positive expectancy set was established showed:

(1) a greater percentage of students being placed on the job earlier

(2) a smaller percentage of students being fired

(3) a lower rate of absenteeism

(4) a smaller percentage of unexcused absences

It should be emphasized that with the exception of item (4), the findings were not significant at the desired level of probability (.05). The results should therefore be

considered as suggestive evidence relating to the basic hypothesis of the research.

The findings, however, do not tell us whether or not the impact of the expectancy was of a self-fulfilling nature or of a halo effect. A reasonable case can be made for both interpretations. The data examined in the next section; turnover, absenteeism, lateness and performance on the job should help clarify the issue. If expectancy effects are observed in the work environment we must conclude that a self-fulfilling prophecy was operating. In that event, a halo effect interpretation would not be appropriate since the new decision-makers (i.e., supervisors and other employees on the job) would not have been aware of the "outstanding potential" designation for some of the graduates.

(2) Inter-personal Expectancy Effects on the Job

In accordance with usual Bank policy, approximately 30-45 working days after a CSTC graduate is placed on a job, a blank Supplemental Performance Appraisal Report (Appendix D) is sent to her supervisor for completion. This report requests information on the absence and lateness record of the employee and on her performance as perceived by the supervisor.

A total of 101 trainees were graduated from the CSTC and were placed in jobs throughout the Bank (25 experimental,

76 control). Performance Appraisal reports were received for 79 of that number (17 experimental, 62 control). Reports were not received on the remaining 22 because:

(a) 13 terminated prior to completion of the form

(b) 9 of the supervisors did not respond, despite repeated efforts, to the request for the report.

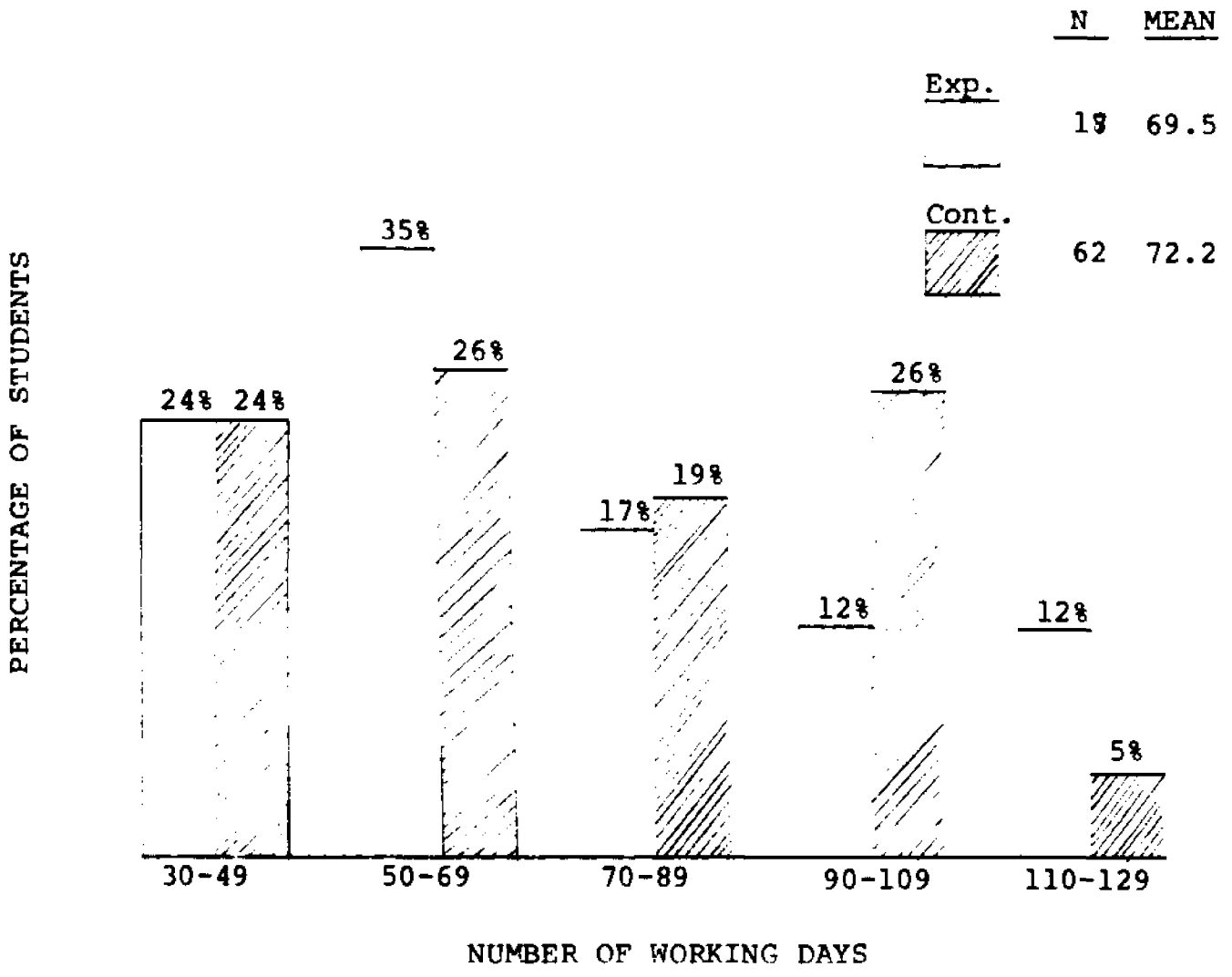
In several cases, the reports received were incomplete (ratings, absence or lateness information omitted) resulting in varying "N's" among the performance data.

Because of procedural delays in sending out and/or completing the appraisal forms, the length of time covered by the reports varied considerably. Since performance data is likely to be a function of time on the job, it was necessary to compare the experimental and control groups for significant differences along the "time covered" dimension. Figure 2 shows the percentage distributions of working days covered by the appraisal reports for the experimental and control groups. A test to determine whether the two distributions were significantly different indicated that they were not (Kolmogorov-Smirnov test).

Turnover Rates. Since turnover statistics are a function of time on the job (i.e., will increase with time), it was necessary to establish a data recording procedure which would eliminate any bias associated with the widely varying

FIGURE 2

PERCENTAGE DISTRIBUTION OF WORKING DAYS COVERED BY THE PERFORMANCE REPORTS



placement dates. Thus, for the purpose of this analysis, terminations were recorded only if they occurred within 90 working days after placement on the job.

A total of 16 (15.8 percent) of the CSTC graduates terminated from their jobs within 90 days of placement. Twelve were control group employees and four were from the experimental group. Table 14 presents a breakdown of the total figures by treatment group and turnover type.

As seen, there are no major differences between the experimental and control groups in their total and voluntary turnover rates. As in the CSTC turnover statistics (Table 11), the involuntary termination comparison favors the experimental group (zero percent as compared to 5.2 percent for the controls). However, since the percentage difference does not approach statistical significance, (.24 level of probability), the result can only be considered as suggestive.

These findings provide no real support to the hypothesis that the experimental group would demonstrate significantly better turnover rates on the job than would their control counterparts. Furthermore, the lack of statistically significant results in the involuntary termination rates implies that the difference in the rates of involuntary terminations while at the CSTC (see page 66) was probably the result of a "halo" effect and not a self-fulfilling prophecy.

TABLE 14

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS
FOR TURNOVER RATES DURING FIRST 90 DAYS ON THE JOB

	EXPERIMENTAL GROUP		CONTROL GROUP			
	N	% OF TOTAL N (25)	N	% OF TOTAL N (76)	z	p
TOTAL TURNOVER	4	16.0	12	15.7	.03	.98
INVOLUNTARY TURNOVER	0	0.0	4	5.2	1.17	.24
VOLUNTARY TURNOVER	4	16.0	8	10.5	.73	.47

Absenteeism and Lateness Rates. Table 15 presents the results of an analysis of the absence and lateness data. The percentages of latenesses were almost identical for both groups - approximately 3 percent. Comparisons of the absence results shows that the experimental students were not in attendance 6.2 percent of the time whereas the controls were absent at a rate of 4.4 percent. A test for the significance of the difference between these two percentages indicates a probability value of .01. This result reverses the trend shown when the trainees were at the CSTC; there the absence rates favored the experimental group 3.9 to 4.5 percent (.17 probability level).

The shift in the trend of the attendance results was primarily an outcome of increased on-the-job absenteeism among the experimental group (3.9 to 6.2 percent; significant at .01 level). The increase could be interpreted as a negative reaction to the new working environment which did not provide the "special treatment" given by the instructors. Lacking this support, the students behaved contrary to earlier teaching and hence, the higher absence rate.

A simpler explanation is to ascribe the low absence rate during training to "purposeful" actions by the CSTC administrators and instructors (halo effect actions). An omitted absence here, a forgotten absence there could easily have led to better attendance results. If this was the

TABLE 15

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR ABSENCE AND LATENESS RATES ON THE JOB

	EXPERIMENTAL	CONTROL	z	p
TOTAL NUMBER OF DAYS ABSENT / TOTAL NUMBER DAYS POSSIBLE TO BE ABSENT	.0618	.0435	2.90	.01
TOTAL NUMBER OF DAYS LATE / TOTAL NUMBER OF DAYS POSSIBLE TO BE LATE	.0334	.0309	.44	.65

case, we would expect, as we found, an increase in on-the-job absence rates for the experimental group and no change among the controls (control group rates were 4.5 and 4.4 percent; not-significant).

Since the Performance Appraisal Report does not provide information on whether or not the employees were paid for the days they were absent (excused absence), it was not possible to determine if the significant result of the Training Center experience (Table 13) was repeated in the job situation.

In summary, the on-the-job attendance and lateness data appears to support the view that the expectancy effects observed during training were of a "halo" nature.

Performance Ratings. The Supplemental Performance Appraisal Report provides the following data pertaining to the employee's work performance.

(1) Supervisory Ratings on Job Performance, Behavior and Attitude. Using a five-point scale, the supervisor rates the employee along several dimensions in each of the three areas (see Appendix D). The point values of the specific ratings in each area are individually summed and then individually divided by the number of rated dimensions to determine the Job Performance,

Behavior and Attitude ratings. For example, if four of the dimensions in the attitude area were rated three and four were rated five, the attitude rating for that individual would be four. (4 x 3 = 12, 4 x 5 = 20, 20 + 12 = 32, 32 divided by 8 is 4)

(2) General Performance Evaluation. This is a one item overall rating on the same five point scale mentioned in (1) (see lower, left corner of the Appraisal Report, Appendix D). The supervisor supposedly weighs the dimensions comprising the Job Performance, Behavior and Attitude categories for importance and, on that basis, determines what the overall rating should be.

A median test was used to identify the differences between the groups in regard to the Job Performance, Behavior and Attitude ratings. For each rating, the median score was determined for the combined group (experimental and control). Then both sets of scores were independently dichotomized at that combined median and the data was cast in a 2 x 2 contingency table. A Chi-Square test was then employed to examine the significance of the data.

Tables 16, 17 and 18 show the results of this analysis. In all three instances, the experimental and control groups were almost identical in the proportion of their scores

TABLE 16

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR DISTRIBUTION OF JOB PERFORMANCE RATINGS FALLING BELOW AND ABOVE THE COMBINED MEDIAN RATING.

	EXP.	CONT.	TOTAL
NUMBER OF RATINGS BELOW THE MEDIAN	9	30	39
NUMBER OF RATINGS ABOVE THE MEDIAN	8	31	39
	17	61	78

COMBINED MEDIAN RATING = 3.3; X^2 = NOT SIGNIFICANT

TABLE 17

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR DISTRIBUTION OF BEHAVIOR RATINGS FALLING BELOW AND ABOVE THE COMBINED MEDIAN RATING.

	EXP.	CONT.	TOTAL
NUMBER OF RATINGS BELOW THE MEDIAN	9	29	38
NUMBER OF RATINGS ABOVE THE MEDIAN	8	31	39
	17	60	77

COMBINED MEDIAN RATING = 3.6; X^2 = NOT SIGNIFICANT

TABLE 19

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR DISTRIBUTION OF ATTITUDE RATINGS FALLING BELOW AND ABOVE THE COMBINED MEDIAN RATING.

	EXP.	CONT.	TOTAL
NUMBER OF RATINGS BELOW THE MEDIAN	9	27	36
NUMBER OF RATINGS ABOVE THE MEDIAN	8	29	37
	17	56	73

COMBINED MEDIAN RATING = 3.4; X^2 = NOT SIGNIFICANT

falling above and below the combined median. There was obviously no difference between the two groups in the distribution of Job Performance, Behavior and Attitude ratings.

The General Performance Evaluation data was examined somewhat differently. The proportion of experimental employees with above average ratings (rated 4 or 5) were compared to their counterparts among the controls. A similar comparison was performed for those employees who were rated below average (rated 1 or 2). The results of this analysis (Table 19) clearly show no meaningful difference between the groups in the percentage of above average or below average rated employees.

In summary, the performance of the two groups on the job, as indicated by supervisory ratings, appears to be indistinguishable. The experimental group did no better nor any worse than did the employees for whom no expectancy was established.

(3) Conclusion

Based upon the results, it is concluded that a self-fulfilling prophecy was not demonstrated in this study. There was no substantial evidence to show that the interpersonal expectancy had a self-fulfilling effect upon learning achievement, job performance, absenteeism, punctuality or turnover. Thus, the hypotheses of the

TABLE 19

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS FOR PROPORTION OF GENERAL EVALUATION RATINGS FALLING BELOW AND ABOVE THE AVERAGE "3" RATING.

	EXPERIMENTAL		CONTROL			
	N	% OF TOTAL N (17)	N	% OF TOTAL N (62)	z	p
RATINGS BELOW 3 (BELOW AVERAGE)	1	.06	4	.07	.09	.93
RATINGS ABOVE 3 (ABOVE AVERAGE)	8	.47	26	.42	.38	.70

research must be rejected.

The few statistically significant and near significant findings which are consistent with the hypotheses appear to be a direct outcome of a halo effect which may have resulted from the administrators' and instructors' knowledge that certain students had "outstanding" potential.

DISCUSSION

The general purpose of this research was to evaluate and extend Rosenthal's finding that interpersonal expectations can be self-fulfilling. Since the results do not indicate that the purpose was accomplished, attention should next be given to the question "Why?" - - "Why did we fail to demonstrate a self-fulfilling prophecy?" This chapter addresses that question.

Validity of Rosenthal's Findings

As indicated earlier, there is a growing body of literature (Barber and Silver, 1968; Thorndike, 1968; Snow, 1969) which criticizes Rosenthal's handling and presenting of data associated with expectancy effects experiments (in the laboratory and in the classroom). Taken together, these critical reviews seriously question the validity of Rosenthal's work.

One objective of the current study was to provide an independent (of Rosenthal) test of the validity of the self-fulfilling prophecy. If evidence could be produced which was similar to or consistent with Rosenthal's results then his work and theories, despite the criticism, would gain credibility. The actual outcome was, of course, entirely non-supportive of Rosenthal's findings. No evidence of a self-fulfilling prophecy was found. It would appear, therefore, on the basis of this experiment, that there is some justification to question the validity of Rosenthal's results. Additional validation studies, especially direct

replications of Rosenthal's experiment, are needed to shed further light on the reality of the self-fulfilling prophecy.

Validity of the Current Findings

The appeal of Rosenthal's thesis is especially enhanced by the fact that several of his "successful" studies took place in real-life settings - notably the classroom (Rosenthal and Jacobson, 1968A; Rosenthal and Evans, 1968; Conn, Edwards, Rosenthal and Crowne, 1968). The study described here was also performed in a real-life environment. While it is not my intention to debate the merits and weaknesses of real-life (as compared to the laboratory) experimentation, the reader should be aware of the constraints faced in this study, since they may lead to questions about the validity of the results.

The experimenter, you may recall, played a relatively passive role throughout the course of the study. His principal task was to implant an expectation among the instructors about the high potential of selected students. He was not permitted to change or disrupt the training conditions or environment in any major way. He could not, for example, introduce new tests to measure achievement, nor could he establish procedures to insure a consistent testing procedure for all students. In short, he could not establish the degree of control usually associated with experimental research.

The requirement to leave the Canal Street system and its components intact bears directly on the validity

of the results. If the system failed to generate reliable and valid data, then the results must be questioned. In this regard, the following limitations of the Canal Street system (and ultimately the study) were noted during the course of the experiment.

(1) Achievement tests. There is no question that the standard achievement tests employed were generally related to the English and mathematics curricula of the CSTC. In measuring learning achievement, however, it is more precise to have criterion instruments that are directly relevant to the behavioral objectives of the curriculum. Since the instructor teaches to those objectives, any differential classroom treatment of students is more likely to show up in examinations which directly test those objectives.

Although a more definitive or appropriate measure of classroom achievement could have been employed, there is no real quarrel with the standard tests actually used. If the expectancy set was operating, there would at least have been suggestive evidence in the data.

(2) Quality of the data. The integrity of the grade-point scoring system depends upon accurate,

consistent and bias-free record-keeping on the part of the instructors. If classroom test data (not standard achievement tests) are recorded and maintained in the same way by all of the instructors, then the grade-point data can be used with a degree of confidence. Since experimenter control over the maintenance of data was lacking, we can never be certain as to differences among the instructors in their record-keeping and/or data reporting practices. Despite expected administrative staff assurances, the reliability of the grade-point data must be questioned.

Similarly, lack of experimenter control over the recording of absence and lateness data duri..g training increases the likelihood of accidental or purposeful omissions, deletions, corrections, etc. While there is no direct evidence to substantiate a claim of intentional or unintentional mis-recording of data, it is a distinct possibility when considering the ease by which it can be done (no checks on administrative staff), the heavy work load of the administrators, and the large number of trainees at the center (anywhere from 120 to 250 trainees at a given time).

Interpretation of on-the-job absence and lateness data are additionally complicated by the fact that dozens of different supervisors (administrators) are involved in recording the data. It can be assumed with confidence that there are differences among the supervisors in their administrative abilities (i.e., record-keeping) and, more important, in the criteria they employ in determining when lateness or absences occur. For example, one supervisor will record lateness exactly at the start of the working day, another will provide an allowance of 15 minutes, another an hour, another will never record latenesses for official disposition. Similarly, some supervisors always record absences, some record absences occasionally, and some never record absences.

On-the-job performance data is also affected by individual differences among the supervisory staff. Despite the guidelines and instructions provided by the bank it is reasonable to assume that different criteria and standards are employed when judging the job performance of an employee. "Four" rated employees of one supervisor may very well be equivalent to the "3" rated employees of another supervisor, and so on. The bias associated with subjective assessments of performance are very well known.

One positive aspect of the on-the-job results (absence, lateness, performance) is that there appears to be no systematic bias in the data. The placement of CSTC graduates in jobs was conducted without any knowledge or reference to the supervisory practices of the department. Thus, we can assume that the "random" placement of trainees resulted in "equivalent" exposure to varied supervisory practices among the control and experimental groups.

(3) Data Collection Procedures. It was previously stated that "it was not possible for E (without uncovering the true intent of his study) to exercise any control over the testing procedures" (page 41). This lack of control resulted in several testing instances where arbitrary decisions were made by the administrative staff to omit certain tests. The ultimate impact of those decisions was an uncomfortably small number of test score entries in the various sub-group cells. This was particularly true in the analysis of the effect of early data returns.

The statistical analyses of grade-point scores were also subject to small "N's" due to inadequate control over data collection procedures. The data

from one entire class (15 trainees) was excluded from the analysis because of irregularities in the computation of scores that went unnoticed by the administrative staff.

The limitations of the CSTC system discussed above have apparently influenced the conduct and results of the study. The magnitude and direction of that influence, however, is not known and can only be resolved through further study.

Success of the Expectancy Manipulation

Another plausible explanation for the inability of this research to replicate Rosenthal's results is addressed by the question: "Was the expectancy set actually established?" If the expectancy manipulation was not successful and the instructional staff did not accept the notion of "outstanding potential" students, then, of course, the findings are meaningless.

Pages 35 through 39 describe, in considerable detail, the elaborate procedure utilized in establishing and verifying the expectancy set. As stated there, only "three of the thirteen teachers interviewed were suspicious of the true intent of the study. - - - - - "The other ten instructors seemed to have a positive

attitude towards the study as indicated by their acceptance of the "outstanding potential" concept and willingness to discuss personality and behavior characteristics of the "outstanding" group." (page 39)

Other evidence of the successful implantation of the expectation comes from the administrative staff of the Center. Throughout the course of the study they unknowingly acted as the communication link between E and the instructors. Any grumblings or concerns among the instructors relating to the "outstanding potential" students would likely have come to their attention and they, in turn, would have informed E. On only one occasion was E so advised - two of the three "doubting" instructors had voiced their concern to the Director of the Center.

Additional support for the view that the expectancy manipulation succeeded is found in the results of the study itself. Indications were that a halo effect was probably operating in favor of the experimental students in connection with time spent in training, involuntary terminations from the Center, and unexcused absences during training. If such a halo effect did, indeed, occur, it would mean that the teachers' and administrators' actions were directly influenced by the statement that certain students had "outstanding potential". In other words, the expectancy manipulation was successful.

If we accept the finding that only two English and

one mathematics instructor were not successfully influenced by the manipulation, the next question is "What impact did the three have on the results?" Since all of the classes had one or the other English teacher, all of the English related test data and grade-point scores are obviously irrelevant. Mathematics test data and mathematics grade-point scores for the classes taught by the "doubting" mathematics instructor are also irrelevant.

The remaining dependent measures are either specific to the other instructors (i.e., mathematics test scores and grade-point scores in mathematics, office procedures, office machines and typing) or are of a general non-instructor specific nature (turnover, absenteeism, lateness, time spent in training). The only influence that the three dissidents could have had on these variables is through communication of their doubts to the other instructors. There is no evidence to suggest that such communication ever took place, or if it did, that it had any effect. None of the 10 "believers", for example, ever mentioned to E or E's confederate that three of the instructors doubted the purpose of the study. Further, the administrative staff of the Center gave no indication that the instructors often discussed the project among themselves. From all indications, it would appear that the expectancy manipulation was successful for 10 of the 13 teachers involved in the study.

Environment and Population Differences

Almost all previous studies of teacher expectancy effects have focused on the elementary school teacher in predominately white schools. This research attempted to extend the self-fulfilling theory to an industrial training environment where the students were adult, disadvantaged and of a minority population (primarily blacks and Puerto Ricans). The lack of success in attempting to confirm the hypothesis generated by the earlier research findings may very well be a result of the different learning environments and populations involved. The following observations and conjectures are offered in support of that possibility.

(1) Time available. The industrial trainer, generally speaking, has a shorter period of time available to accomplish his objectives than does the school teacher. He therefore would have less overall time to interact, communicate and shape the behavior of the "outstanding" student.

(2) Continuity of interaction. Ideal conditions for the actualization of a learning achievement self-fulfilling prophecy would appear to include a relatively continuous day-long interaction between "expecter and expectee". The elementary schools generally provide for the one class-one teacher requirement. Because of the wide

scope of the curriculum, such continuity of interaction is usually absent in industrial training programs. For example, in this study, the trainees met with several teachers during the course of a day, each of whom focused on different teaching objectives. It is possible that the relatively brief daily exposures to the students were insufficient for the shaping of the desired expectancy effects to occur.

(3) Teaching Objectives. In a very broad sense, the fundamental objective in a school system is to provide the student with the skills and knowledge required for "life". In the industrial world, the objective is much more precise - - it is to provide the trainee with the skills and knowledge required to do a specific job. It is hypothesized here that the operational frame of reference which arises out of the industrial need to train for a job may inhibit the successful occurence of a self-fulfilling prophecy.

In industry, the instructor's job is dependent upon his ability to turn out trained employees. If he consistently fails to do so, he is at fault, not the students. This is unlike the school situation where poor performance on

achievement or intelligence measures is a fact of life and is generally attributed to the student's (not the teacher's) ability. With this in mind, consider the industrial trainer who is told about "outstanding" trainees. What advantage would it be to him to consciously (or unconsciously) shape the performance of a few students when he has to have all of them prepared and up to standard? He cannot afford to focus on the special students because of the possible negative impact on the performance of the others (i.e., they might not neet performance standards). In my opinion, the inherent conflict in complying with the pressures of an expectancy situation tends to reduce the probability that a self-fulfilling prophecy can be demonstrated in an industrial learning situation.

In the school environment, special treatment of selected students would not present any problem to the teacher. It is common practice in the schools to give individualized attention - - the expectancy situation demands no more than that. Since it is consistent with the regular teaching procedures

to comply with the behavioral requirements of the expectancy, demonstrations of the self-fulfilling prophecy in that environment are more likely to occur.

(4) Student Age and Social Class. One of Rosenthal's findings was that the self-fulfilling prophecy was primarily observed among children of the younger grades (Rosenthal and Jacobson, 1968A). An obvious implication of that finding is that age of student influences the probability that a self-fulfilling prophecy occurs. The older the student, the less likely that the expectancy will be self-fulfilling. Poor learning habits, previous experiences in school, problems at home, low motivation level for school, are just a few of the factors associated with increasing age that may interfere with the expectancy set.

If, as in this study, the students are adults and black or Puerto Rican, then the list of possible interfering variables should be increased by social class associated factors such as; language difficulties, biases of white middle class instructors (ten of the 13 instructors were white), health problems, level of aspiration, history of academic failure, etc.

In view of the above considerations, it is reasonable to hypothesize that the failure to replicate may have resulted from differences in the teaching environments and the student populations.

Concluding Remarks

Various alternative explanations have been presented to account for the failure to validate and extend the theory of a self-fulfilling prophecy. Chief among these is the view that the validity of Rosenthal's work is suspect. The findings of this study obviously support the growing number of critics who question Rosenthal's methods, results and conclusions. Additional studies, especially direct replications of Rosenthal's work, are needed to determine the "truth" of the self-fulfilling prophecy.

REFERENCES

Adair, J. G., and Epstein, J. S. Verbal cues in the mediation of experimenter bias. Psychological Reports, 1968, 22, 1045 - 1053.

Barber, T. X. and Silver, M. J. Fact, fiction and the experimenter bias effect. Psychological Bulletin Monograph, December 1968, 70, 1 - 29.

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

Beecher, H. K. Pain: one mystery solved. Science, 1966, 151, 840 - 841.

Beecher, H. K. Surgery as placebo. Journal of the American Medical Association, 1961, 176, 1102 - 1107.

Bernstein, B. Social Structure, language and learning. In Passow, A. H., Goldberg, M., and Tannenbaum, A. J. (eds.) Education of the Disadvantaged. New York: Holt, Rinehart and Winston 1967, 225 - 244.

Burnham, J. R. Effects of experimenter's expectancies on children's ability to learn to swim. Unpublished master's thesis, Purdue University, 1968.

Claiborn, W. L. An investigation of the relationship between teacher expectancy, teacher behavior and pupil performance. Unpublished doctoral dissertation, Syracuse University, 1968.

Clark, K. B. Dark Ghetto. New York: Harper and Row, 1965.

Coleman, H. A. The relationship of socio-economic status to the performance of junior high school students. Journal of Experimental Education, 1940, 9, 61 - 63.

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

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

Curry, R. L. The effect of socio-economic status on the scholastic achievement of sixth-grade children. British Journal of Educational Psychology, 1962, 32, 46 - 49.

Deutsch, M. The Disadvantaged Child . New York: Basic Books 1967.

Deutsch, M., and Brown, B. Social influences in Negro-white intelligence differences. Journal of Social Issues, 1964, 20, 24 - 25.

Doddy, H. H. The progress of the Negro in higher education: 1950 - 1960. Journal of Negro Education, 1963, 32, 485 - 492.

Dreger, R. M., and Miller, K. S. Comparative psychological studies of negroes and whites in the United States: 1959 - 1965. Psychological Bulletin Monograph Supplement, September, 1968, 70.

Friedman, N., Kurland, D and Rosenthal, R. Experimenter behavior as an unintended determinant of experimental results. Journal of Projective Techniques and Personality Assessment, 1965, 29, 479 - 490.

Haase, W. The role of socioeconomic class in examiner bias. In Riessman, F., Cohen, J., and Pearl, A. (eds.) Mental Health of the Poor. New York: Free Press, 1964, 241 - 247

Hollingshead, A.B., and Redlich, F. Social Class and Mental Illness. New York: Wiley and Sons, 1958.

Larrabee, L. L., and Kleinsasser, L. D. The effect of experimenter bias on WISC performance. Unpublished manuscript. Psychological Associates, St. Louis, 1967.

Marwit, S. J., and Marcia, J. E. Tester bias and response to projective instruments. Journal of Consulting Psychology, 1967, 31, 253 - 258.

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

Merton, R. K. Social Theory and Social Structure. Glencoe, Illinois: Free Press, 1949. Pp 179 - 195.

Miller, D. R., and Swanson, G. E. Inner Conflict and Defense, New York: Henry Holt, 1960.

Milner, E. A study of the relationship between reading readiness in grades on school children and patterns of parent - child interaction. Child Development, 1951, 22, 95 - 112.

Pasamanick, B., and Knobloch, H. The contribution of some organic factors to school retardation in Negro children. Journal of Negro Education, 1958, 27, 4 - 9.

Pitt, C. C. V. An experimental study of the effects of teachers' knowledge or incorrect knowledge of pupil IQ's on teachers' attitudes and practices and pupils' attitudes and achievement. Unpublished doctoral dissertation, Columbia University, 1956.

Putnam, C. Letters to the editors. Perspectives in Biology and Medicine, 1963, 7, 138 - 139.

Rice, S. A. Contagious bias in the interview: a methodological note. American Journal of Sociology, 1929, 35, 420 - 423.

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

Rosenthal, R. Teacher expectations. Chapter to appear in G. S. Lesser (Ed.) Psychology and the Educational Process. Glenview, Illinois: Scott, Foresman. 1969 (expected).

Rosenthal, R., and Anderson, D.F. Teacher behavior and the mediation of teacher expectancy effects. Unpublished data, Harvard University 1969.

Rosenthal, R.,and Evans, J. Unpublished data reported by Rosenthal (1969), Harvard University, 1968.

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

Rosenthal, R., and Fode, K. L. Three experiments in experimenter bias. Psychological Reports , 1963, 12, 491 - 511. (B)

Rosenthal, R., and Jacobson, L. Pygmalian in the classroom: Teacher Expectation and Pupils' Intellectual Development. New York: Holt, Rinehart and Winston, Inc., 1968. (A)

Rosenthal, R., and Jacobson, L.F. Teacher expectations for the disadvantaged. Scientific American, 1968, 218, 19 - 23. (B)

Rosenthal, R., Kohn, P., Greenfield, P. M., and Carota, N. Experimenters' hypothesis - confirmation and mood as determinants of experimental results. Perceptual and Motor Skills, 1965, 20, 1237 - 1252.

Rosenthal, R., and Lawson, R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. Journal of Psychiatric Research, 1964, 2, 61 - 72.

Rosenthal R., Persinger, G. W., Vikan-Kline, L., and Fode, K.L. The effect of early data returns on data subsequently obtained by outcome-biased experimenters. Sociometry, 1963, 26, 487 - 498.

Sexton, P. C. Education and Income: Inequalities of Opportunity in our Public Schools, New York: Viking, 1961.

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

Snow, R. E. Unfinished Pygmalian. Contemporary Psychology, 1969, 14, 197 - 199.

Tanser, H. A. The Settlement of Negroes in Kent County, Ontario, and a Study of the Mental Capacity of their Descendents. Chatham, Ontario, 1939.

Thorndike, R. L. Review of Pygmalian in the classroom. American Educational Research Journal, 1968, 5, 708 - 711.

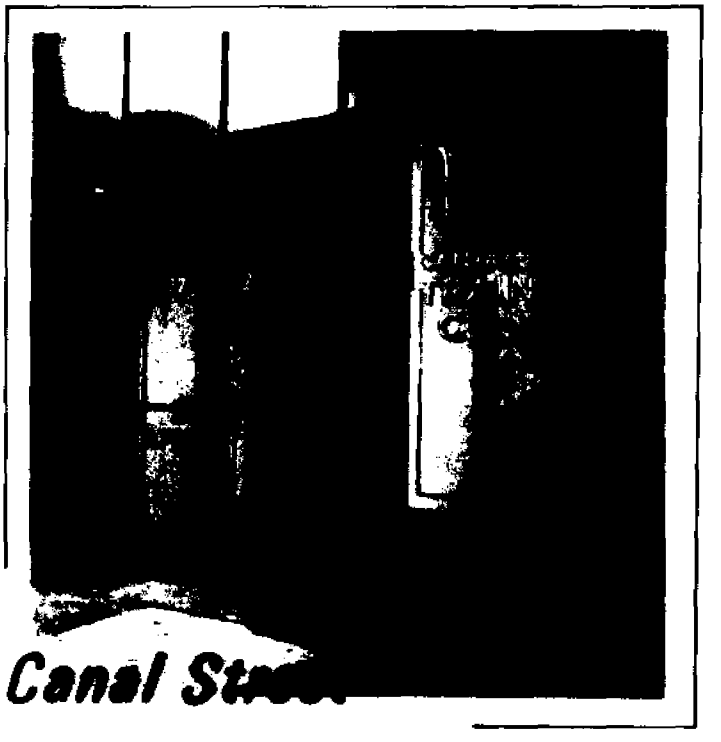
Wilson, A. B. Social stratification and academic achievement. In Passow, A. H. (ed.) Education in Depressed Areas, New York: Teachers College Press, 1963, 217 - 235.

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

APPENDIX A

DESCRIPTION OF CANAL STREET TRAINING CENTER

First National City Bank



Canal Street Training Center

WHAT IS THE CENTER ?

The Canal Street Training Center is a school operated by the Bank to train men and women for entry level jobs as clerks or typists within the Bank.

WHY IS THE BANK OPERATING THIS CENTER ?

At the request of the President of the United States, employers in the nation's 50 largest cities have assisted in an effort to employ and train, at Federal Government expense, the disadvantaged of our land. For FNCB, this Center provides a pool of well-trained employees to meet the Bank's continuing personnel needs.

First National City Bank

Canal Street Training Center

WHAT IS THE CENTER ?

The Canal Street Training Center is a school operated by the Bank to train men and women for entry level jobs as clerks or typists within the Bank.

WHY IS THE BANK OPERATING THIS CENTER ?

At the request of the President of the United States, employers in the nation's 50 largest cities have assisted in an effort to employ and train, at Federal Government expense, the disadvantaged of our land. For FNCB, this Center provides a pool of well-trained employees to meet the Bank's continuing personnel needs.

HOW ARE THE GRADUATES PLACED IN THE BANK ?

A graduate of the Canal Street Training Center is made available for interview and selection by any group within the Bank. Many Supervisors have begun to use the Training Center as their regular pool from which to select well-trained replacements.

HOW ARE THE GRADUATES DOING ON THE JOB ?

No formal graduation ceremonies are held because the members of most classes have already completed training and are well on their respective ways to becoming successful employees in various sections of the Bank. The early reports indicate that the Training Center is producing an employee who is enthusiastic and productive.

WHERE IS THE TRAINING CENTER ?

The Center is located at 277 Canal Street, at the corner of Broadway and Canal. The Bank leased the second and third floors and over 16,000 square feet have been divided into class rooms.

WHY IS THIS PROGRAM IMPORTANT?

For the disadvantaged of our nation, this program provides an opportunity to gain skills and become productive citizens.

For First National City Bank, this program provides a supply of well-trained employees to meet the unceasing need generated by our growing business.

HOW ARE THE GRADUATES PLACED IN THE BANK ?

A graduate of the Canal Street Training Center is made available for interview and selection by any group within the Bank. Many Supervisors have begun to use the Training Center as their regular pool from which to select well-trained replacements.

HOW ARE THE GRADUATES DOING ON THE JOB ?

No formal graduation ceremonies are held because the members of most classes have already completed training and are well on their respective ways to becoming successful employees in various sections of the Bank. The early reports indicate that the Training Center is producing an employee who is enthusiastic and productive.

WHERE IS THE TRAINING CENTER ?

The Center is located at 277 Canal Street, at the corner of Broadway and Canal. The Bank leased the second and third floors and over 16,000 square feet have been divided into class rooms.

WHY IS THIS PROGRAM IMPORTANT?

For the disadvantaged of our nation, this program provides an opportunity to gain skills and become productive citizens.

For First National City Bank, this program provides a supply of well-trained employees to meet the unceasing need generated by our growing business.

WHO IS BEING TRAINED ?

Under the Department of Labor's Manpower Act, those who fail to complete school are classified as disadvantaged in the labor market. In addition, those who are unemployed or underemployed, those who are handicapped, those who have sub-standard incomes, or those who are members of a minority group are judged to be disadvantaged.

In FNCB's Canal Street Training Center, the majority of the students are women who are heads of households. Over fifty-five per cent were on welfare at the start of their training. Sixty-five per cent were school dropouts. Ninety per cent had been unsuccessfully employed. However, over eighty per cent had received no prior job training. The relationship between limited education and training and past job difficulties may be significant.

WHEN DID THE PROGRAM BEGIN ?

The program began in July of 1968 and will continue through November of 1969. During this period, seven hundred men and women will have received training and will have been placed on jobs in the Bank.

WHAT KIND OF TRAINING IS GIVEN ?

The training is designed to develop entry level employees for placement within the Bank. Twenty new trainees enter the program with each new class. For the first eight weeks they receive basic education in arithmetic, reading and verbal expression, plus preliminary training in office practices and office skills. The program also includes training in work habits and office conduct. Counseling is provided the trainees on personal, financial and work related problems.

After evaluation, some are selected to specialize in typing training (12 more weeks); while the others continue in general clerical and office machine training (8 weeks). When a student has demonstrated that he is qualified, he is transferred to productive work.

APPENDIX B

GRADE POINT SCORING SHEET

Last Name _____ First Name _____ Group____ Entered_____

15th Wk_____

TYPING

	Units	Score
Lessons 1 - 30	3	____
Lessons 31 - 49	2	____
Lessons 50 - 64	1	____
Lessons 76 -80 and rough drafts	1	____
Lessons 81 -90	2	____
Lessons 101 - 110 and bank forms, memos	1	____
Typing Tests for Bus. Straight Copy	1	____
Manuscript	1	____
Letters	1	____
Timed Writings:		
8th wk ____ 9th wk ____		
10th wk ____ 11th week ____	1	____
12th wk ____ 13th wk ____		
14th wk ____ 15th wk ____		
16th wk ____	1	____
Special Project	3	____
Special Timed Writings	2	____
Total - Typing		____

ENGLISH

Topic	Units	Score
Verbs	2	____
Spelling	2	____
Punctuation	2	____
Agreement	2	____
Adjectives & Adverbs	2	____
Pronouns	2	____
Vocabulary	3	____
Bonus Points		
Letters	1	____
Compositions	1	____
Book Reviews	3	____
Total - English		____

Check-point scores:

Each grade
90 - 100 6 points
80 - 90 5 points
70 - 80 4 points

Multiply the points times the value of the units to get score for each topic.

You must have 350 points to be placed. Extra points win extra dividends.

Summary of Scores

Typing ____

English ____

Mathematics ____

Office Machines ____

Office Procedures ____

TOTAL ____

Last Name _____ First Name _____

Group ---- Entered _____

15th Wk _____

MATHEMATICS

Topic	Units	Score
Whole Numbers	6	
Decimals	3	
Fractions	3	
Arithmetic of the Typewriter	1	
Percentage and Discount	2	
Bonus Points		
Measurement	1	
Interest	2	
Salaries, Taxes, Payrolls	3	
Total - Math		

OFFICE MACHINES

10-key	Units	Score
Machine Principles	1	
Test 1	1	
2	2	
3	2	
Check runs (1)	1	
(2)	2	
Full-bank		
Machine Principles	1	
Test 1	1	
2	2	
3	2	
Check runs (1)	1	
(2)	2	
Bonus Points		
Calculators	2	
Total OM		

OFFICE PROCEDURES

Topic	Units	Score
Filing - Rules	1	
Pr. Appl'n	1	
Numeric	1	
Cards	1	
Telephone Usage - Tech.	1	
Office work	1	
Classroom	1	
The Banking System	2	
Bank Terminology	1	
Bank Forms	1	
Tube System & Mailing	1	
Duplicating	1	
Interviews	2	
Bonus Points		
_____	5	
_____		
Total - O.P.		

APPENDIX C

STANDARDIZED TEST BATTERY

PLEASE NOTE:

Appendix C, "Numerical Ability Test", (c) 1947 by The Psychological Corporation; "Fundamental Achievement Series", (c) 1966 by The Psychological Corporation; "Vocabulary", (c) 1968 by First National City Bank; "Numbers" and "Office Terms", (c) 1956 by Industrial Psychology, Inc. not microfilmed at request of author. Available for consultation at City University of New York Library.

UNIVERSITY MICROFILMS.

APPENDIX D

PERFORMANCE APPRAISAL FORM

D-1

PER 239 12-68

SUPPLEMENTAL PERFORMANCE APPRAISAL REPORT

INSTRUCTIONS — *COMPLETE SECTIONS A THROUGH E WITH REFERENCE TO THE EMPLOYEE NAMED AT THE BOTTOM RIGHT OF THE PAGE. RETURN THE COMPLETED FORM IN THE ENVELOPE PROVIDED.*

SECTION A – EMPLOYEE INFORMATION

DEPARTMENT		LOCATION		POSITION CODE/PAY GRADE	DATE ASSIGNED	WORKING HOURS
OVERTIME REQUIRED ☐ YES ☐ NO	NO. WORK DAYS	DAYS ABSENT	DAYS LATE	YEARLY SALARY ➧	STARTING	CURRENT

JOB DUTIES

OFFICE MACHINES EMPLOYEE OPERATES

LAST JOB PERFORMANCE MEASUREMENT (E.G. WORDS PER MIN.) AND DATE

SECTION B.C.D. PERFORMANCE RATINGS

EVALUATE EACH ITEM BASED UPON OBSERVED PERFORMANCE IN PRESENT POSITION (USE QUANTITATIVE DATA WHEREVER POSSIBLE TO ARRIVE AT AN APPROPRIATE EVALUATION)

KEY: N A – NOT APPLICABLE NOT AVAILABLE.
1 – UNACCEPTABLE – NEEDS CONSIDERABLE IMPROVEMENT.
2 – CONDITIONALLY ACCEPTABLE-NEEDS TO IMPROVE.
3 – GOOD – TOTALLY SATISFACTORY PERFORMANCE.
4 – VERY GOOD – REGULARLY PERFORMS ABOVE EXPECTED LEVEL.
5 – EXCELLENT-CONSISTENTLY PERFORMS AT AN OUTSTANDING LEVEL

SECTION B – JOB PERFORMANCE	NA	1	2	3	4	5
VOLUME OF WORK HANDLED						
ACCURACY						
ATTENTION TO DETAIL						
INITIATIVE						
EXPEDITING WORK UNDER PRESSURE						
ORGANIZING OWN WORK						
ABILITY TO WORK WITHOUT SUPERVISION						
WILLINGNESS AND ABILITY TO WORK WITH OTHERS						
ADAPTABILITY TO NEW ASSIGNMENTS						
TECHNICAL KNOWLEDGE OF OPERATIONS						
CONCENTRATION						
USE OF SUPPLIES						
CARE OF EQUIPMENT						
CLEANING UP						
COMMUNICATIONS						
OVERALL QUALITY OF TYPING OUTPUT						
OVERALL QUALITY OF MACHINE OPERATION						
OVERALL QUALITY OF CLERICAL WORK						
OTHER:						

SECTION C – BEHAVIOR	NA	1	2	3	4	5
DRESS						
GENERAL GROOMING						
GENERAL MANNERS						
COURTESY TO SUPERVISOR						
COURTESY TO OTHER EMPLOYEES						
COURTESY TO PEOPLE FROM OUTSIDE DEPT						
ATTENDANCE						
PUNCTUALITY						
OTHER						

SECTION D – ATTITUDE						
TOWARD FIRST NATIONAL CITY BANK						
TOWARD JOB DUTIES						
TOWARD WORK LOCATION						
TOWARD ADDITIONAL EDUCATION						
TOWARD OTHER ETHNIC GROUPS						
TOWARD SELF						
TOWARD BANK RULES AND REGULATIONS						
OTHER						

SECTION E – OVERALL APPRAISAL

DO YOU FEEL THIS EMPLOYEE WAS WELL TRAINED FOR THE JOB HE OR SHE IS CURRENTLY PERFORMING? YES NO PLEASE EXPLAIN.

POTENTIAL FOR PROMOTION TOO EARLY TO JUDGE

RECOMMENDED DEVELOPMENT PROGRAM TO STRENGTHEN WEAK AREAS OF PERFORMANCE TOO EARLY TO JUDGE

GENERAL PERFORMANCE EVALUATION

(HEAVILY WEIGH THOSE FACTORS MOST IMPORTANT TO PERFORMANCE IN THE POSITION)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

GENERAL COMMENTS

SUPERVISOR'S NAME			PERSONNEL NO.	DO NOT MARK
SUPERVISOR'S TITLE	PHONE EXT.	DATE	EMPLOYEE'S NAME	

APPENDIX E

ABSTRACT OF ALLEGED RESEARCH STUDY

RESEARCH ABSTRACT

James Coleridge and Mark Samuels. The use of tests as a learning predictor. California Journal of Educational Research, 1969, 22, 140 - 148.

This research was part of an extensive effort to investigate the effectiveness of a basic skills (reading and arithmetic) adult education program conducted by the Department of Education in San Diego, California. The ninety-five students in the program were primarily (90%) minority group adults (Mexicans and Negroes) referred by local businesses for remedial aid. The program was conducted for 16 weeks at a centrally located high school during the evenings from 7 - 10 P.M.

Prior to the start of the program, each student was required to complete a battery of achievement and personality tests. These tests included the: California Achievement Tests (Advanced); California Basic Skills Tests (Intermediate); Fundamental Achievement Series (Form A); Edwards Personal Preference Schedule and; Gordon Personal Profile.

Correlational analyses of performance on the test battery to achievement in the program as determined by teacher grades were non-significant for all but the Fundamental Achievement Series. An r of .51 was obtained for the Numerical Test and an r of .46 for the Verbal Test. Both of these correlations were significant at p. 01.

While pointing out the need for further evidence, the authors nevertheless conclude that the "Fundamental Achievement Series appears to have considerable value in predicting who the outstanding students in an adult population will be". (page 147)

APPENDIX F

SAMPLE MEMORANDUM SENT TO INSTRUCTORS COMMUNICATING THE NAMES OF "OUTSTANDING POTENTIAL" STUDENTS

MEMORANDUM TO: Instructional Staff
Canal Street Training Center

The following students of the class beginning August 13, 1969 (Group 4A) have been identified as having outstanding potential to succeed in training. Please note your records accordingly.

Ronald J. Freeman
Hattie Jennett
Rosa Sierra
Estela Vazquez
Ethel Wright

Howard S. Mase, Director
Training Evaluation

August 15, 1969

cc: Dr. Norman Willard, Jr., V.P.
Ralph Roberts, Director

APPENDIX G

MEMORANDUM SENT TO INSTRUCTORS SUMMARIZING THE NAMES OF ALL "OUTSTANDING POTENTIAL" STUDENTS

COPY SENT TO ALL INSTRUCTORS ON ATTACHED LIST

To:

For your general information the following list represents all of the students selected in connection with my special study.

Class IA	Class 33	Class 2A
Martha L. Barriera	Norris H. Brown	Marcia H. Brisset
Sheila Ilano Boone	Ivy Folkes	Eleanor Hayes
Vera Jean Caleb	Joyce N. Jones	Linda Hillman
Lydia Baudino	Deborah McNeill	Mildred Robertson
Florence Wheaton		Patricia Scroggins
		Hazel Woods

Class 3A	Class 4A	Class 5A
Sadie Brinson	Ronald J. Freeman	Catherine Bonner
Maria Miranda	Hattie Jennett	Janice Chestnut
Virginia Moore	Rosa Sierra	Shirley Gordon
Grade Roach	Estela Vazquez	Annie Grant
Geneva Smith	Ethel Wright	Adel Rodriguez

In weeks to come I will be meeting with each of you individually to discuss your observations in this project.

Howard S. Mase

Howard S. Mase, Director
Training Evaluation

September 3, 1969

APPENDIX H

INTERVIEW INSTRUMENT

INSTRUCTOR OBSERVATIONS OF SPECIAL STUDENTS - **CSTC**

Instructor:_____

Subject:_____

Date of Interview_____

(1) **Who were the "outstanding potential" students in your classes? (Names)**

(2) Observations re: Similarity in:

(a) personality traits (e.g., intelligence, achievers, etc.)

(b) attitudes (e.g., friendly, hostile, etc.)

(c) classroom behavior (e.g., outspoken, leader, etc.)

(3) What kinds of educational practices/techniques would work best with "high potential" students?

(4) Any other comments re: the "outstanding potential" students?