AN OBSERVATIONAL STUDY OF TEACHERS' EXPECTANCY EFFECTS AND THEIR MEDIATING MECHANISMS ON STUDENTS IN PHYSICAL EDUCATION ACTIVITY CLASSES

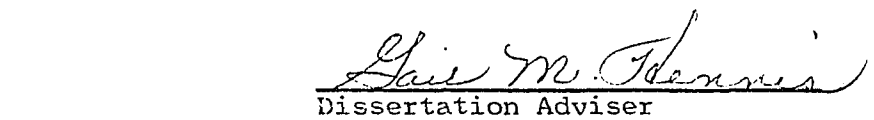
by

Patricia Barbara Crowe

A Dissertation Submitted to
The Faculty of the Graduate School of
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Greensboro
1977

Approved by



Dissertation Adviser

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Dissertation Adviser _____

Committee Members _____

_____

_____

_____

March 15, 1977
Date of Acceptance by Committee

CROWE, PATRICIA BARBARA. An Observational Study of Teachers' Expectancy Effects and Their Mediating Mechanisms on Students in Physical Education Activity Classes. (1977)
Directed by: Dr. Gail Hennis. Pp. 190

The purpose of this study was to identify specific and differential teacher behaviors that affect student behavior based on Rosenthal's Four Factor Theory (1974). Rosenthal's four variables included Climate, Feedback, Output and Input. An additional variable, Touch, was included as a fifth factor. These variables were used to identify teachers' differential treatment of students according to the teachers' expectations, and to identify students' differential responses according to the teachers' expectations (high or low).

Four different physical education activity classes were selected for observational study. Teachers were asked to rank their students (total group in each class) in order of their physical achievement or skill potential. The rankings were used as the criterion measure of the teachers' expectations for their students' performance in physical education. Three judges, trained in the use of the Brophy and Good Interaction Analysis System (1969), observed 96 (24 students from each class) junior high school students on six separate days within a two-week period. Forty-eight of the students were designated as high achievers and 48 of the students were designated as low achievers.

Summary sheets were developed for separate tabulations of the coded observations. Twenty-four frequency measures and 32 percentage measures were derived from the coding. Analyses of Variance

were performed on the five variables to assess the effects of teacher expectations and class, and their interactions on the obtained rankings, and to determine the effect of five different variables on high and low achievers. The data yielded the following results:

1. A significant difference was found indicating that the designated high achievers were given more opportunities to respond and were asked more questions by the teachers than were the designated low achievers.

2. Climate. A significant difference was found which showed that teachers treated the designated high achievers more warmly than they treated the designated low achievers.

3. Feedback. There was a significant difference in the amount of affirmation and praise given indicating that teachers directed more evaluative comments to the high achievers. There was minimal evidence to show that teachers gave other kinds of feedback more to their high achievers than to their low achievers. The non-significant statements outweighed the number of significant statements, and strong support could not be given for the Feedback Factor.

4. Output. A significant difference was evident indicating that the designated high achievers received more attention and were given more opportunities to respond. There was no significant difference to show that teachers gave more reinforcement to the designated high achievers than they did to the designated low achievers. As a result, strong support could not be given for the Output Factor.

5. Input. There was no significant evidence to suggest that teachers taught more new material to the designated high achievers than they did to the designated low achievers.

6. Touch. There was no significant evidence to show that teachers touched their designated high achievers, nor did teachers exhibit any more Climate, Feedback, Output or Input when they touched the designated high or low achievers.

It was concluded that stronger support was needed to show that teachers communicate their expectations to their students through differential teacher behaviors. Additional evidence is needed in the field of physical education to further corroborate Rosenthal's Four Factor Theory.

ACKNOWLEDGEMENTS

The writer wishes to express appreciation to Dr. Gail Hennis, Graduate Advisor, and Dr. William Powers, Consultant for data analysis, for their assistance in this study.

Special thanks are extended to the following graduate students: Jackie Daley, Ellen Greaves, and Brenda Meese, who served as judges, and Claudia Ferguson and Betty Harris who added needed support and assistance.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Sociologists, medical doctors, and behavioral scientists have referred to the concept of the self-fulfilling prophecy in economics, industry, areas of international tensions, and institutions of medicine and education (Allport, 1950; Clark, 1963; McClelland and Winter, 1969; Merton, 1957; Passow, 1963; Rosenthal, 1974). Dr. Robert Rosenthal (1968) has indicated that people do what is expected of them and we are able, to some extent, to predict behavior because of certain norms or expectations imposed by society.

The premise stated for the expectancy effect is that people will behave as they believe they are expected to behave (Rosenthal, 1974). Many times we have preconceived ideas about people. One explanation for this is that we may know about a person's past behavior, so we proceed to predict future behavior. Another reason could be attributed to a person's appearance or background. Behavior may be judged or predicted by what a person wears, the color of his skin, his ethnic background, his geographical location, his I.Q., and/or his intellectual demonstration. Rosenthal and Jacobson (1968) have stated that an interpersonal self-fulfilling prophecy is one which shows ". . . how one person's expectation for another person's behavior can quite unwittingly become a more accurate prediction simply for its having been made" (p. viii).

Merton (1957) developed the concept of the self-fulfilling prophecy and implied that this theory not only had very important implications in the field of economics and industry, but could be a crucial determinant for minority groups and race relations. If low expectations and standards can reinforce a sense of failure among culturally disadvantaged people, then these groups become victims of the self-fulfilling prophecy since failure can possibly reinforce inferior feelings (pp. 421-436). Another theorist, Allport (1950), suggested that the expectancy of armed conflict could be communicated to opponents who react to this expectation, and, in turn, the initiator's expectation is strengthened and confirmed. The above illustrations indicate that the reinforcement is a type of feedback loop system. McClelland has done considerable research in the area of economics and has demonstrated that achievement motivation training can produce economic change. Economic growth can be accelerated by means of psychological education. In their book, Motivating Economic Achievement, McClelland and Winter (1969) summarized their research as follows:

> What seems to be essential is that the man develop a strong faith in himself as an origin or agent of change. If he believes in himself, if he is motivated to change things then he is an expert on how to carry out change . . . . It seems far more effective to convince a man directly that he can accomplish what he wants and then let him find ways to do this. (p. 349)

Other expectancy effects have been evident in the healing profession, by the hypnotist, the psychotherapist, and the physician. Rosenthal (1968, 1974) has cited studies in which the physician has communicated his enthusiasm and confidence for a

new drug as compared to that same physician's conservative report on further research of that same drug. Rosenthal (1968) has stated that "This phenomenon, like the operation of placebo effects in general, can be partially understood in terms of the healer's expectation for the efficacy of the preparation" (p. 16). The physician, after considerable research and confirmed reports, may find that he has doubts about the drug's effects and somehow communicates this to his patient.

In the areas of psychotherapy and hypnosis, beliefs and expectations of the therapists concerning their patients have induced appropriate responses to the therapists' own expectations. Rosenthal (1968) has said that ". . . briefly put, prospective patients are given psychotherapy lessons, they learn what to expect and what will be expected of them" (p. 13).

Educational theorists, throughout the years, have discussed the positive and negative outcomes of teachers' expectations on pupils' intellectual development and potential. Clark (1963), when writing about the effects of student-teacher relationships regarding aspirations and achievements, referred to the self-fulfilling prophecy by saying:

> If a teacher believes that a child is incapable of being educated, it is likely that this belief will in some way be communicated to the child in one or more of the many forms of contacts inherent in the teacher-pupil relationship. (p. 183)

Goldberg (1963) indicated that if a child from a low socio-economic and culturally deprived background is treated as if he is uneducable because he has a low test score, he may very well

become uneducable and the low score is reinforced. She emphasized that conviction by saying:

> It is highly probable that the lower class poor achiever is viewed more negatively than the middle class achiever, therefore teacher responses may play a stronger part than expected in the development of self concept to the extent that the child's feelings of acceptance by the teacher raises his estimate of himself. (p. 96)

Rosenthal (1974) stated that "until recently the evidence for the hypothesis of the self-fulfilling prophecy has been observational or correlational rather than experimental" (p. 1). Much attention was given to the disadvantaged child, pointing to the fact that these particular children were victims of teachers' educational self-fulfilling prophecies (Clark, 1963; Goldberg, 1963). However, there was no experimental research to support this persuasive and obvious theory. The question to be answered was the nature of the intervening variables responsible for the effect.

In an effort to find answers and to document evidence that a teacher's expectations or prophecies could make some difference in either her evaluation of her students or her students' actual performance, Dr. Rosenthal (1963, 1964, 1968, 1969) undertook extensive studies in both the laboratory (animal and human subjects) and in the educational classroom.

Rosenthal's work addressed a major social problem and as a result, the "expect-effect" phenomenon precipitated a number of replications as well as critical evaluations (Thorndike, 1968; Jensen, 1969; Clairborn, 1969; Snow, 1969). In addition, other studies were conducted relating to manipulated or existing teacher

expectations with some studies focusing on the learner and some studies focusing on teacher-pupil interaction (Goldsmith, 1971; Jose & Cody, 1971).

Since 1970, investigators (Brophy & Good, 1970; Rothbart, 1971; Rubovitz & Mayer, 1971) have made an attempt at explaining the Pygmalion Effect rather than placing their emphasis on replications of the effect. For the past 15 years, the research on expectancy effects has accumulated and there is enough evidence to support Rosenthal's theory of the self-fulfilling prophecy, namely, that the expectancy effect does indeed exist. Rosenthal has stated, however, that it is now time to accumulate more evidence on how this effect has been operating in classrooms and other related situations. Rosenthal (1974) has stressed the necessity of continued research by saying that, "many studies are needed, both laboratory and field, by different workers, in different centers of research" (p. 24).

Statement of the Problem

Need for the Study

After a research of the available literature and after receiving confirmation from Dr. Rosenthal at Harvard University (1975), the writer found only one study of teacher expectations in the area of motor performance. Burnham's (1968) results of teacher expectations in a swimming class revealed the existence of expectancy effects. However, his study was not concerned with the identification of the behavioral mechanisms by which teacher expectations affect student behavior. In effect, there have been no studies in

physical education identifying the mechanisms or processes by which teacher expectations are communicated to pupils.

It is the writer's contention that if expectancy effects occur in the classroom and mediating influences affect student behavior, they also may be evident in the gymnasium even though the nature of the activity is different. If, as Rosenthal (1968) has suggested, the teacher influences a student's self-image and the student depends on the teacher for encouragement and reinforcement, then it may be important to look at differential teacher behavior. The writer believes that there is a need to study this problem by observing student-teacher interaction for the identification of the operation of the Pygmalion Effect and the intervening factors responsible for the effect.

In 1973, Rosenthal (1974) reviewed and summarized all of the studies of the self-fulfilling prophecy. He focused on those studies which revealed evidence of the mediation of self-fulfilling expectations operating in classrooms, offices, and factories. From his preliminary evidence, Rosenthal devised a four factor "theory" on the mediation of self-fulfilling expectations. His contention was that teachers, counselors, and supervisors who expect superior performance from their charges treat them differently than expected inferior performers in four particular ways:

1. Climate: Teachers appear to create a warmer socio-emotional climate for their "special" students in the following ways: Smile, wink, establish and maintain eye contact, pat on back, place hands on student, raise eyebrows and smile, give sign or

any gesture of approval, indicate friendliness, support and understanding, have pleasant-sounding voice.

2. Feedback: Teachers appear to give to their "special" students more differentiated feedback as to how these students have been performing. More attention is given and more active teaching occurs with special students. Both Climate and Feedback involve differential teacher warmth toward students of whom more versus less is expected. However, if a teacher shows warmth and gives praise specifically in response to a correct response, or helps to correct a response (by giving clues or rephrasing), or asks for further information, then the Feedback factor would be operating.

3. Input: Teachers appear to teach more material and more difficult material to their "special" students. As compared to the Feedback factor where active teaching occurs (praise for correct response, correcting incorrect response, or giving clues), the distinction between Feedback and Input is in the amount of new material and more difficult material taught to students of whom more is expected.

4. Output: Teachers appear to give their "special" students greater opportunities for responding. Skill and competence is encouraged by the teacher and greater demands may be imposed by the teacher on students of whom more is expected. These demands and opportunities can take form in the following ways: giving students more time to respond or perform, calling more often on those students, or asking those students to perform more difficult tasks or answer more difficult questions.

Purpose of the Study

It was the purpose of this study to investigate and identify specific and differential teacher behaviors that affect student behavior based on Rosenthal's Four Factor Theory (1974, pp. 14-24). Specific purposes included the following:

1. Identification of teachers' differential treatment of students according to the teachers' expectations (high and low).

2. Identification of students' different responses according to the teachers' expectations (high and low).

3. Explanation of additional factor, namely, touch as another possible identifying mechanism responsible for the Pygmalion Effect.

Major Hypothesis

The hypothesis to be tested in this study was based on the assumption that a teacher's expectation could make some difference in evaluation of his/her students.

Sub Hypothesis

1. Expectancy effects occur in physical education activity classes.

2. Teachers treat students differently according to the teachers' expectations (high and low expectations).

   A. Teachers who expect superior achievement from their students treat them differently as compared to their low or inferior achievers in four particular ways: climate (warmth), feedback, input, and output (Rosenthal, 1974, pp. 14-24).

B. Teachers touch their high-achieving students more than their low-achieving students.

C. Teachers who touch their high-achieving students more also exhibit more warmth, feedback, input, and output.

3. Teachers' expectations influence student behavior according to the teachers' predicted evaluations of the students' behavior.

Assumptions Underlying the Study

The following assumptions governed this study.

1. People will behave as they believe they are expected to behave. This behavior may be manifested in a positive or negative manner. A person who holds an expectation for another person's behavior will communicate this expectancy to that person, thereby influencing him to respond in accordance with the expectations (Rosenthal, 1974).

2. Specific mechanisms by which teacher expectations are communicated to students can be identified through an observational interaction analysis system.

3. In physical education activities, the physical manipulation of the student occurs frequently. Teachers help students in physical education activities by correcting improper grips, adjusting stances, moving them through particular patterns of movement, and keeping them in balanced positions.

Scope of the Study

This study was limited by the following factors:

1. The study was conducted in one junior high school in Greensboro, North Carolina.

2. The subjects were seventh-, eighth-, and ninth-grade junior high students.

3. Four different physical education activity classes, taught by four different teachers, were used in this study.

4. The time period for training and investigation extended over a two-month period from approximately September to mid-November.

5. A three-member observation team was trained by the investigator. The team consisted of three graduate students in physical education enrolled at the University of North Carolina at Greensboro.

6. There were six separate observations for each of the four classes. These observations extended over a two-week period.

7. Twenty-four subjects were selected from each class for observation. Each observer was responsible for eight students.

8. A total of 96 subjects were selected for this study. Each observer was responsible for 32 dyads.

9. No attempt was made to investigate the personal likes or dislikes of students to being touched.

10. No attention was given to whether touching was a desirable or undesirable technique for teachers to use in physical education classes.

Definitions

The following terms used in this study are defined as follows:

1. Self-Fulfilling Prophecy: People will behave as they believe they are expected to behave. This behavior may be manifested in a positive or negative manner (Rosenthal, 1968).

2. Expectancy Effects, Pygmalion Effect, Expect-Effect Phenomenon: Other common usages for the term, Self-Fulfilling Prophecy.

3. Mediating Influences or Mechanisms: How the person who holds an expectation for another person's behavior communicates the expectations to that person, thereby influencing him to respond in accordance with the expectancy (Rosenthal, 1974).

4. Intervening Variables: Another common usage for the above term.

5. General Touch: A teacher moves toward her student and pats student on back, squeezes student's arm, puts arm around student's shoulder, or touches any part of the student's body indicating warmth or a feeling of friendliness.

6. Assistant Touch: The teacher moves to manually manipulate student in a particular movement or touches a student to help with a task.

7. Procedural Touch: Teacher moves toward student to place student in line or move student to a particular spot on the floor.

8. Incidental Touch: Teacher touches student while demonstrating a skill. Teacher stops class to explain a skill and may touch a student.

9. Behavior Touch: Teacher touches a student when giving student a warning or criticism or praise of the student's behavior.

10. Dyadic Interaction: In the study of dyadic interactions, the individual student or the teacher-child dyad becomes the unit of analysis rather than the class as a group (Brophy & Good, 1969).

CHAPTER II

REVIEW OF LITERATURE

The review of literature examines the experimental evidence that has accumulated as a result of two main influences: Dr. Rosenthal's early research on experimenter effects and artifacts in behavioral research, and the observational claims from educational theorists that disadvantaged children were the victims of teachers' educational self-fulfilling prophecies. The literature is divided into five sections: (a) Experimental Studies in the Laboratory, (b) Experimental Studies Outside the Laboratory, (c) The Mechanisms Operating on Teacher Expectations, (d) Rosenthal's Four Factor Theory, and (e) Objections and Critical Reviews of Rosenthal's Work.

Experimenter Effects in Behavioral Research

In their book, Artifact in Behavioral Research, Rosenthal and Rosnow (1969) pointed out two major effects which the experimenter could have upon the findings of his research. The first effect is not interactional and consequently does not affect responses of the subjects. Observer effects are those in which the experimenter might record errors, whether biased or not, in the direction of his hypothesis. The second major effect operates by affecting the subjects' responses directly. This type is concerned with the expectancy of the researcher relating to his

hypothesis and results of research. Biosocial effects, psychosocial effects, situations and subjects may all affect the expectancy of the experimenter. All of these artifacts such as environment, different personalities, age and sex differences, and acquaintance of subjects are all unintentional but nevertheless, do affect subject responses. The above illustrations do not necessarily affect the subject's treatment condition, but expectancy effects by the experimenter of how his subjects will react do change the function of the treatment condition. Rosenthal (1969) has stated that "the expectancy of the experimenter about the subject's behavior may contribute to a determination of what that behavior will actually be" (p. 196).

A. Experimental Studies in the Laboratory

Rosenthal (1966, 1969) with others (Fode, 1963; Lawson, 1964); Adair & Epstein, 1968) conducted a number of experiments in animal and human behavior to demonstrate how the attributes of the experimenter could affect subject responses. The answers supported the theoretical question of the effect of the interpersonal self-fulfilling prophecy posited by educators.

The first study was undertaken by Rosenthal and Fode (1963) in the early part of the 1960's. They investigated the proposition that one person's expectation for another person's behavior could come to serve as a self-fulfilling prophecy. They asked 10 graduate students in Experimental Psychology to be the experimenters and 10 undergraduates in Introductory Psychology to be the experimenters' subjects. The subjects were to identify

photographs on the basis of success and failure shown in the faces of the photos. The experimenters were given bogus information about a previous experiment indicating significant results and were asked to duplicate these results. Half of the experimenters were told that their subjects would rate the photos as successful and the other half of the experimenters were told that their subjects would rate the photos as not successful (expectancy induction). The results showed that expectations of the experimenters affected their subjects' responses. Higher photo ratings were obtained when experimenters expected higher photo ratings compared to the experimenters who did not expect successful ratings. A replication by Adair and Epstein (1968) was undertaken to find out why or by what means this expectancy effect had occurred. Tape recordings were taken of the experimenters' identical instructions to their subjects. Then the recordings were played to the subjects. Again, when success was expected, the subjects rated the photos as successful which seemed to indicate that the subjects' responses were affected by the tone of voice (tape recordings). The self-fulfilling effects of the experimenters' expectations were demonstrated.

One well-known experiment was the case of Clever Hans (1966). Clever Hans, a brilliant horse, could add, subtract, multiply, divide, spell, and read. The owner of Clever Hans, Mr. Van Osten, said that he never gave cues to the horse. Mr. Van Osten asked others to test the horse's talents. Pfungst (1965) undertook research to discover how Clever Hans operated. He found that

if the questioners were not visible, or if the questioners did not know the answers, the horse could not respond. Pfungst finally discovered (after many observations) that various unintentional cues such as inclination of the questioner's head, raising eyebrows, dilating nostrils, and moving toward and away from the horse at certain times, were the causes of Hans' brilliance. Rosenthal (1966) stated that:

> Hans' questioners, even skeptical ones, expected Hans to give the correct answers . . . . Their expectation was reflected in their unwitting signal to Hans that the time had come for him to end his tapping. The signal cued Hans to stop and the questioner's expectation became the reason for Hans' being, once again, correct. (p. 196)

In other studies, the behavior of rats was investigated to determine the effects of experimenter expectancy. In Rosenthal and Fode's (1963) study, experimenters were told that some rats were maze dull and some were maze bright. The results indicated that animals believed to be brighter showed daily improvement over the rats who were labelled dull. There was no difference between the dull and bright animals, but the experimenters were led to believe otherwise. When both groups of rats responded correctly (ran to the rewarded side of the maze), the rats believed to be brighter ran faster. When the experiment was over, the examiners were asked to rate their rats and to describe their own behavior and feelings about their subjects. The evaluation revealed that the experimenters who thought they had bright rats saw them as pleasant and likeable. They also said that they treated the brighter animals more warmly and gently, they watched them more

carefully, and they were more friendly and enthusiastic toward them.

Another study, conducted by Rosenthal and Lawson (1964) also used rat subjects but utilized the Skinner Box rather than the maze. The Skinner Box was more complex and challenging. Again, supposedly, brighter animals were superior performers because of the experimenter's expectancy of excellent performance. As in the other maze-learning experiment, the experimenters were asked to rate their subjects and their own behavior and attitudes. The experimenters with the brighter rats as subjects treated their subjects more warmly, were more gentle and soothing, and they watched their animals more closely.

In other experiments conducted on human subjects, tasks such as person perception, reaction time, and inkblot tests, have been utilized. In over 90 studies cited by Rosenthal (1966, 1969), the interpersonal self-fulfilling prophecy was demonstrated. Every experimenter who was lead to believe one thing or to expect particular responses received those responses. Unintentional effects were manifested in different ways: facial expressions (positive or negative), auditory cues, and visual cues. According to Rosenthal (1968), it is these unintentional influences that cause these unintended behaviors which then lead subjects to respond as prophesied. He has stated that "Probably neither subject nor experimenter knows just exactly what the unintended communication behavior is and neither do we" (Rosenthal, 1968, p. 30).

B. Experimental Studies Outside the Laboratory

After innumerable studies undertaken with rats, and after observations on other human research studies, Rosenthal (1969) suggested that "Many of the effects of the experimenter including the effects of his expectancy may have considerable generality for other social relationships" (p. 196). From these results, Rosenthal (1966) wondered about the effects of the self-fulfilling prophecy outside the laboratory. He asked the question:

> When the master teacher or school principal believes a junior teacher's pupils to be slow learners, is this belief (well founded or not), likely to accelerate or decelerate these pupils' educational progress? (p. 140)

To emphasize the importance of such phenomena, Rosenthal and Jacobson (1966) attempted a study in 1966 to see if "teachers' expectations of their pupils' ability might, in fact, be a partial determinant of those pupils' ability" (p. 410). Their procedure was basically the same as in the experiments on the effects of the experimenter's expectancy. The study was expanded and in 1968 the complete results were published in a book entitled Pgymalion in the Classroom.

The authors (1968) asked themselves (concerning the advantaged as well as the disadvantaged), "Is there any good evidence that a teacher's expectations or prophecies make any difference in either her evaluations of her pupils or in their actual performance" (p. 54)? They applied the theory of the self-fulfilling prophecy and provided substantial evidence indicating that teachers'

expectations of students' intellectual competence could come to serve as an educational self-fulfilling prophecy.

The study was conducted in a public elementary school located in a lower socio-economic status neighborhood. All children were given a non-verbal test of intelligence; one that would predict intellectual blooming. The teachers were given an explanation of the research and were told that a particular test, to be given to the children, would predict which children were most likely to show an "academic spurt" (Rosenthal & Jacobson, 1968, p. 66). The purported "Harvard Test of Inflected Acquisition" was really "Flanagan's Test of General Ability" (TOGA). Flanagan's test was chosen because it had not been used routinely at the school and teachers were apt to be unfamiliar with this particular test. Eighteen classrooms, three at each of six grade levels, were composed of children with above-average ability, average ability, and below-average ability. Approximately 20% of the children were randomly selected to form the experimental group. Each teacher was given the names of the children from her class who were in the experimental condition. The teachers were told that these children had scored high on the test for intellectual blooming and they should show remarkable gains in intellectual competence during the following eight months of school. The only difference between the experimental group and the control group was in the minds of the teacher. After eight months, the children were retested with the same IQ tests (TOGA). The children of the experimental group showed only a slight gain in the verbal IQ over

the control group. But in the reasoning IQ test and the total IQ test, the experimental group gained considerably more than did the control group.

Rosenthal and Jacobson also noted that the children in the experimental group were perceived by the teachers as interesting, curious, appealing, and affectionate. The authors (1968) concluded that ". . . it would seem that when children who are expected to grow intellectually do so, they are considerably benefitted in other ways as well" (p. 108). On the other hand when those children who are not expected to excel, do so, they are assessed as showing undesirable behavior. The authors (1968) stated:

> If a child is to show intellectual gain it seems to be better for his real or perceived intellectual vitality and for his real or perceived mental health if his teacher has been expecting him to grow intellectually. It appears worthwhile to investigate further the proposition that there may be hazards to unpredicted intellectual growth. (p. 118)

Rosenthal and Jacobson's conclusions of the Pygmalion study precipitated a number of replications. Researchers, interested in teacher expectancy effects, were eager to replicate these findings for the purpose of disproving Rosenthal and Jacobson's theory, or to shore up and firm the existing data.

In the Rosenthal and Jacobson study (1968), girls had greater gains in intellectual blooming than boys. Among the boys, those who were expected to bloom gained less than the children of the control group. To check this finding, the Pygmalion experiment was repeated by Evans and Rosenthal (1969). Children were from

middle-class backgrounds and the results were in the opposite direction. This time, the boys showed the benefits of favorable teacher expectations and the girls who had been expected to bloom intellectually gained less in reasoning IQ than girls in the control group. All of the children showed substantial gains in IQ. Evans and Rosenthal (1969) stated:

> These results while they suggest the potentially powerful effects of teacher expectations also indicate the probable complexity of those effects as a function of pupils' sex, social class, and, as time will no doubt show, other variables as well. (p. 263)

Another study was conducted by Conn, Edwards, Rosenthal and Crowne (1968) at an East Coast school of upper middle-class pupils. Both boys and girls who were expected to bloom intellectually showed increased gains in reasoning IQ over those shown by boys and girls of the control group, and the magnitude of the expectancy effect favored the girls slightly. According to the authors (1968):

> It was of considerable theoretical interest to find that greater benefits of favorable teacher expectations accrued to those children who were more accurate in judging the emotional tone expressed in an adult female's voice. (pp. 33-34)

The findings also suggested that vocal cues may play a part in the covert communication of interpersonal expectancies.

Clairborn (1969), critical of Rosenthal and Jacobsons' methods and data, used 12 first-grade classrooms (four groups of three grades each). Clairborn's purpose was to: (a) observe teacher pupil interaction after the teacher received bogus information about the intellectual potential of her pupils, and (b) to replicate Rosenthal's findings that teacher expectancies may

bring about an effect in her students' intellectual performance, if the teacher perceives them as "special" (p. 317). The criterion measures were IQ gain (using Flander's Test of General Ability) and observation of teacher-pupil interaction. Two months later, after retest, the special pupils showed no relative gains. There were no clear changes in observed teacher-pupil interaction. It was concluded that the evidence for bias effects in school remains equivocal. Though Clairborn stated that he replicated as much as possible, there were many differences between his and Rosenthal and Jacobson's study.

J. Jose and Cody (1971) partially replicated the Rosenthal and Jacobson study using the TOGA test and a standard achievement test. They too were interested in whether a teacher's behavior changed toward her students after receiving false information, as well as how student IQ scores and achievement tests changed. The investigators believed that little attention or thought was given to what actually occurred between the teacher and student after the teacher was given an expectancy induction. An interaction analysis scale was used to observe any changes in teacher behavior after establishment of expectancy. After 16 weeks, post-measures of IQ gain, changes in reading and arithmetic achievement, and teacher-pupil interaction were obtained. The authors found no significant differences in teacher behavior. The expectancy had little effect on the teacher's overt behavior.

J. S. Goldsmith and E. Fry (1971) reported a partial replication showing no significant expectancy effects on IQ gains.

Subjects were 112 experimental and 112 control high-school students. TOGA was administered as well as the Sequential Tests of Educational Progress (STEP). Expectancy lists (lists of bloomers) were given to the teachers, and they were reminded several times during the five-month experimental period of the list of special students. Postmeasures were taken five months later showing that the criterion measures (gains in IQ scores and gains in scores of the STEP test) were not significant.

Another partial replication was conducted by S. Kester (1969). He used the Standard Achievement Test (SAT), and IQ test (Otis-Lennon), and an attitude test as pupil pretests and posttests. The posttests were given nine weeks after the pretests. In addition, Kester was interested in teacher-pupil interaction. One hundred and fifty seventh-grade pupils were randomly assigned to experimental and control groups and teachers were given the names of the supposedly bright students. The teachers were told that these students would be observed as part of a study on classroom behavior of bright students. In both groups, teachers were observed for the first seven weeks to determine positive verbal and nonverbal interactions (verbal praise) directed at the student. The author was unable to find any significant expectancy effects on the pupil measures. However, teacher observation showed that teachers talked more to bright students and spent more time with bright students who showed more positive behavior toward the teacher. In general, teachers showed favorable interest and were more supportive toward the supposedly "bright" students.

Flowers (1968) used fictitious ability groupings to learn about teacher expectancy and student performance. Two seventh-grade classes were selected from each of two different schools. One class in each school was labelled as a high-ability group, but the teacher knew nothing about the arbitrary nature of the grouping in the control class. At the end of the school year, all of the students were retested on reading and arithmetic and for IQ. The group labelled as high-ability performed better than the control group in one school in reading and arithmetic, but the effect was not dramatic. There were no differences in IQ between the groups. There was no difference between the groups in reading and arithmetic in the other school, but the thought to be high ability group had gained in IQ points. In this particular study, the results were inconsistent.

In the only study measuring physical performance, Burnham (1968) used as subjects boys and girls aged seven to 14 attending a summer camp for the disadvantaged. None of the children could swim at the beginning of the camp session. The camp staff was led to believe that half of the children showed unusual potential for learning to swim as judged from a battery of psychological tests. Children were assigned randomly to the high potential group. All of the children were tested using the Standard Red Cross Beginner's Test. The results indicated that the "high-potential" children showed greater improvement in swimming ability than did the children who were not expected to show increased improvement.

C. Mechanisms Operating on Teacher Expectations

Other studies, prompted again by the Rosenthal-Jacobson study, as well as the replicated investigations, were conducted for the purpose of looking further at manipulated or existing teacher expectations. The area that was being explored was not the existing expectancy effects per se, but the exploration of the mediating influences or mechanisms by which teacher expectancies affect students. The important question to be answered by these investigators was the knowledge of events intervening between the inducement of the expectancy and the administration of the posttest.

Meichenbaum, Bowers and Ross (1969) examined the effects of expectancy instructions on the academic and classroom behavior of institutionalized adolescent juveniles over a one-month period. Because of the length of the experimental period (two weeks), changes in intellectual development were not expected. Therefore, the author did not administer an IQ test as a criterion measure. Fourteen girls, all taught by four different teachers, were the subjects. Six of the subjects were identified as "potential or late intellectual bloomers" (p. 307). All four teachers were given the same expectancy induction. Teachers were asked to rate the subjects' intellectual and academic potential (on the basis of exams and classroom behavior) on a seven-point scale from minimum to maximum potential. Three "good" pupils and three "poor" pupils were then selected as late bloomers on the basis of a previous test predicting academic

potential. Criterion measures were objective and subjective exams, and grades given by the teachers. Academic performance was measured on a pre-post basis comparing grades that the subjects received one month before the study began with the final grades that they received in June. Teachers' classroom behavior was observed before and during the experimental period. The subjects' behavior was also observed during the second and fifth week of the study. A significant expectancy effect was evident on the objective exams but not on the subjective exams and the results of the study supported the evidence that expectancy instructions to teachers about pupils' academic potential significantly modifies pupils' behavior. The expectancy instructions were significantly effective in modifying the subjects' academic and classroom behavior even when the teachers had a low prior expectancy of the pupils' intellectual performance. Most significant was the fact that prior expectancy could also influence teacher behavior. The authors stated that "one means of modifying behavior of both teacher and pupil is to modify the teacher's perception or label of the student's academic potential" (p. 315). The results also indicated that the expectancy effect was mediated due to the quality of interaction between teachers and the expectancy group and not the quantity or increased attention shown the expectancy group.

Palardy (1969) investigated teachers' beliefs of first-grade boys' probable success in learning to read. He was interested in determining whether teachers' reported beliefs about first-grade

boys' probable success in reading had any significant effect on the measured achievement in reading that the pupils in their class attained. Of the 42 teachers asked to give their opinions concerning probable success, 10 were chosen: five who thought boys' probable success was equal to girls, and five teachers who indicated that boys' probability of success was lower than girls. The Standard Achievement Test was given in September and then in May to 53 boys and 54 girls in Group A (boys' success equal to girls), and to 58 boys and 51 girls in Group B (boys probable success lower than girls). The results indicated that the boys in Group B scored considerably lower in reading achievement than girls in either group and boys in Group A. The author stated that when teachers naturally believed that boys are less successful than girls, boys achieve less as compared to the boys of the teachers who believed or had positive expectations.

Shrank (1968) investigated the expectancy effects of enlisted airmen at the United States Air Force Academy Preparatory School. One hundred students were randomly assigned to five class sections of mathematics. These five classes were then randomly designated as five different ability groups. Criterion measures were test and course grades. The author was interested in determining whether assigning ability-level labels to randomly grouped math class sections had an effect upon their academic achievement. Neither the instructors nor the students knew which of the sections were randomly grouped or grouped according to ability. There were significant differences for the highest and lowest labelled sections.

Also each ability-level labelled section achieved higher means than the next lower labelled section. The author stated:

> These results indicate that there is definitely a labelling effect present in simulated ability grouping even though the grouping is actually random. It seems possible, indeed, probable, that this effect upon academic achievement is also present and perhaps dominant in actual ability grouping. (pp. 50-52)

A second study by Shrank (1970), similar to the first, failed to show the labelling effect of ability grouping. In the second study, the investigator informed the instructors that the students were not grouped according to ability but assigned randomly. The author indicated that because the experiments were identical in every aspect except for one major difference (telling instructors that ability groups were simulated), "It would not seem to be the pupil's reaction to his teacher's expectation that produces the labelling effect, but rather the teacher's reaction to his own expectation of his pupil's performance" (p. 360).

Seaver (1971) investigated sibling expectancy effects to see if teachers teaching an older and a younger sibling had the same expectations for both siblings. Seaver's subjects were first graders whose older siblings had been first graders at the same school and who had been taught by the same teacher. The experimental group consisted of children whose older siblings were taught by the same teacher and the control group were those children whose older siblings were not taught by the same teacher. In both groups, subjects were defined as having bright older siblings and not-so-bright older siblings on the basis of the older sibling's first grade IQ, Standard Achievement Test scores, and

grade point average. The younger siblings were then compared using six SAT's and grade point average for the first grade. Results indicated that younger siblings showed greater achievement than the control group when the older siblings were considered bright and taught by the same teacher. Younger siblings of poor students showed less academic achievement when the older sibling had been taught by the same teacher. Also, younger siblings did better with new teachers than subjects did with teachers who had their older siblings.

Another study was conducted by Beez (1970) on 60 preschoolers from a summer Head Start program. Each child was taught the meaning of a series of symbols by one teacher. Thirty of the teachers had been led to expect good symbol learning and the other 30 teachers had been led to expect poor symbol learning. Seventy-seven percent of the pupils that were expected to have better performance learned five or more symbols, whereas only 13% of the children expected to have poor performance learned five or more symbols. The pupils' performance was assessed by an experimenter who was unaware of the particular expectancy induction given to the teacher. It was found that the teachers who were given positive expectations taught more symbols than the teachers who were given unfavorable expectations. The difference in teaching effort was very important, as eight or more symbols were taught by 87% of the teachers who were expecting better performance, compared to 13% of the teachers who were expecting poor performance.

Brown (1969) used 10 teacher trainees to tutor eight first graders on a paired associate learning task. Bogus information

on IQ and personality was given to the tutors. Children were tutored to associate states with capitals. The results indicated that teachers attempted to teach more to the allegedly brighter students and taught less to those students believed to be dull. Of the bright students, 45% or 18 students were taught seven or more associations, whereas only 22% or nine students in the supposedly dull group were taught seven or more state-capital associations. Brown's study supported that of Beez.

Rothbart, Dalfen and Barrett (1971) were also interested in explaining the mediating mechanisms of teachers' expectations. They observed the teachers' allocation of time (amount of attention) directed to the high-expectancy and low-expectancy children. The amount of reinforcement given to both groups and the teachers' evaluations of both groups were recorded. Thirteen undergraduate students were asked to lead a discussion in literature with 52 students who were divided into four groups. Each teacher was given four students, two of whom were labelled as bright and two described as lacking in academic potential. Results showed that there were no differences in the amount of reinforcement (positive or negative) toward the subjects in the two groups. Teachers paid more attention to the better students and these students, in turn, responded more than the alleged dull subjects. Teachers also described the brighter students as having greater potential for future success and needing less approval, whereas the low-expectancy students were perceived as having a higher need for approval. The authors did note that although the data

indicated that there was no difference in the amount of positive or negative reinforcement, ". . . it would be premature to conclude that verbal or gestural encouragement did not serve as a medium for transmitting teacher expectations" (p. 53).

As in the Rothbart study, Rubovitz and Maehr (1971) did not attempt to replicate the expectancy effect but, rather, to make an attempt at explaining it. The authors were interested in Rosenthal's (1974) suggested "interaction quality" hypothesis: the kind of teacher behavior that would affect student performance after an expectancy effect. Twenty-six female undergraduates (interested in teaching as a career) and 104 sixth and seventh graders were involved in the study. Each teacher was assigned four students. Teachers were given bogus information and two students were randomly described as gifted students, while the other two students were described as average students. Teacher-pupil interactions were observed and coded. The recorded behaviors were teacher attention, teacher encouragement, teacher elaboration, teacher ignoring, and teacher praise and criticism of students' statements. Two teacher behaviors, attention and praise, were significant. Teachers requested more statements from the gifted students as compared to the regular students, and teachers also praised the gifted students' statements more than the statements of the ordinary students. In contrast to Rothbart's study, there was no significant difference in the total amount of attention paid to either group. The authors concluded with statistical evidence that the expectations influenced the quality of teacher-student interaction.

In a study by Good (1970), four first-grade classrooms were used to assess the opportunity given by the teachers for pupils to respond in class. Good (1970) suggested that "the wheel of opportunity does not operate randomly in the classroom" (p. 193). The teachers' existing expectancies were used as a measure to rank pupil achievement (high, medium, and low). The teachers were told that experimenters were observing pupils to identify behavior characteristics of the pupils associated with distinct levels of achievement and that the teachers' ranking of pupil achievement would guide an observer in viewing and classifying pupil behavior. Significant differences were found among the three groups favoring the high achievers. The high achievers received more response opportunities than did the low achievers.

In another observational study by Brophy and Good (1970), four first-grade teachers were asked to rank their high and low scholastic achievers in their classrooms. As in Good's previous study, the rankings were used as the measure of the teachers' expectations for classroom performance. In each class, three boys and three girls ranked as high and three boys and three girls ranked as low were selected for observational study. Teacher-child observations were recorded on four separate days in each of the four classes. It was found that high achievers initiated significantly more contacts with their teachers than low achievers. The data showed that the teachers consistently favored the high achievers over the low achievers in demanding and reinforcing quality performance. There was no significant difference relating

to total number of responses, but there was a difference in quality in the total pattern of dyadic contacts. The "highs" were more frequently praised when correct and less frequently criticized when incorrect.

Rist (1970) observed a class of black ghetto children for a three-year period (kindergarten through second grade) and found that teacher differential treatment of differently judged children was clearly evident. Rist observed a tracking system that developed early and persisted throughout the three years where children were sorted into groups of promise and no promise. Rist indicated that the general quality of teacher interaction showed evidence of discrimination between groups of favored and nonfavored children.

Rosenthal's Review of Mechanisms Operating on Teacher Expectations

In 1969, Snow reviewed Pygmalion in the Classroom and stated that Pygmalion did not show any evidence adequately identifying the process by which teacher expectations were communicated to pupils. Rosenthal (1968) had also stated in his book that the researchers were unable to identify any specific teacher behaviors that might have caused the dramatic changes in the expectation group. Teachers did not spend more time with the expectancy group and some teachers did not remember some of the children who were identified as potential spurters (pp. 155-156). Rosenthal stated that the teachers treated the experimental group differently from the control group via facial expressions, gestures, and possibly touch.

Rosenthal (1974) viewed over 242 studies (185 in the laboratory and 57 outside the laboratory) of experimenter effects on humans and animals and suggested that "At the time of the Pygmalion experiment, there was considerable evidence that interpersonal self-fulfilling prophecies could occur, at least in laboratory settings" (p. 11). The unexpected finding in the Pygmalion study was the teachers' descriptions of their pupils' behavior. Rosenthal (1968) has cited his own study and others by Shore (1969), Leacock (1969), and Rubovitz (1971) who investigated the proposition that there may be hazards to unpredicted intellectual growth (p. 12). In these studies the evidence clearly indicated negative consequences of students' unexpected intellectual development.

In Shore's (1969) study, teachers who viewed children in a negative light or who had negative expectations of pupils who had performed well, rated those children as lower in personality and adjustment.

The Leacock (1969) study revealed that children scoring high who were not expected to score high were seen more negatively when they exceeded the teacher's expectations, and when children performed as they were expected, teachers viewed them more positively.

In Rubovitz and Maehr's (1971) study of black and white children, the nongifted children (black and white) were not treated very differently in terms of praise and criticism responses by teachers. There was, however, a difference in the

way the "gifted" black and white children were treated. Gifted white children received more praise than criticism. The ratio of praise to criticism decreased for the black "gifted" children.

Rosenthal (1974) stated that:

> Taken together the results of the studies by Rosenthal and Jacobson, by Shore, by Leacock, and by Rubovitz and Maehr suggest rather strongly, that there may indeed be hazards to a child showing unexpected intellectual potential or development. (p. 13)

D. Rosenthal's Four Factor Theory

After looking at the results of the above studies, Rosenthal was prompted to pursue, in depth, the question of the mediation of interpersonal experimenter effects in everyday life situations. The question, he said, which remained to be answered, was ". . . how the person who holds an expectation for another person's behavior communicates this expectation to that person thereby influencing him to respond in accordance with the expectations" (Rosenthal, 1974, p. 14).

In 1973, Rosenthal (1974) devised a Four Factor Theory on the mediation of self-fulfilling expectations. Rosenthal's contention was that teachers, counselors, and supervisors who expected superior performance from their charges treated them differently than inferior performers in four particular ways:

> Climate. Teachers appear to create a warmer socio-emotional climate for their "special" students.
>
> Feedback. Teachers appear to give to their "special" students more differentiated feedback as to how these students have been performing.

Input. Teachers appear to teach more material and more difficult material to their "special" students.

Output. Teachers appear to give their "special" students greater opportunities for responding. (p. 14)

Rosenthal's criteria for using four factors instead of a lesser or greater number of factors was that, ". . . for each factor there must be at least five empirical studies in support and not more than just a small number of results in the opposite direction" (Rosenthal, 1974, p. 24). All of the factors are correlated but each factor is distinct and distinguishable from the other. Rosenthal (1974) attempted to place over 30 studies dealing with the mediation of experimenter effects into each factor group with some of the studies overlapping into two or more factor groups.

Climate

In the Climate Factor Group, Rosenthal (1974) stated that 15 of the studies (industrial, educational, and clinical) supported the hypothesis of the first factor, and two of the studies gave results in the opposite direction. In general, whether in the clinical, educational, or industrial areas, when therapists, teachers, or supervisors believed their subjects to be compatible, brighter, or successful, these patients, students, or workers were treated more warmly than those who were expected to be less compatible, more dull, or less successful (p. 15).

Alpert (1970) found that therapists acted more warmly toward patients whom they thought were specially selected to be compatible in their sessions compared to the control group, who were considered to be less compatible.

In an experiment by Chaiken, Sigler, and Derlega (1972), the teachers were divided into three groups of bright, dull, and control. Teachers were asked to teach a unit on home and family safety. Teachers were told that the bright student had an IQ of 130, the dull student had an IQ of 85, and in the control condition, the teachers were told that IQ score information had been misplaced. Results showed that teachers looked in the eyes of bright students more, smiled more, and nodded their heads more as compared to the full and control students.

Dalton (1969) discovered that in a naturalistic setting (no induced expectancies) that teachers who were asked to divide their children into high, medium, and low groups provided encouragement 50% of the time. However, for the low group, 31% of interactions were positive, whereas for the high group, 73% of the interactions were positive.

In other studies mentioned earlier (Jose & Cody, 1971; Kester, 1969; Meichenbaum, et al., 1969; Rist, 1970; Leacock, 1969), researchers found that teachers expecting children to be brighter or to learn more do treat those special children more warmly than students who are not as bright or who are not expected to learn more.

Feedback

Ten studies were listed as relevant to the second mediating factor, feedback. Eight of the studies supported the hypothesis, one study did not support the hypothesis, and one study yielded results in the opposite direction (Rosenthal, 1974, p. 18).

Compared to the Climate Factor showing warmth toward pupils of whom more versus less is expected, the difference between the two factors is the degree to which warmth or praise is given for a correct or desired response or the appropriate feedback given for an incorrect response, such as asking more questions or being critical of that response. Rosenthal noted that even though a teacher directly criticizes an incorrect response from a child, the teacher still can be warm toward that child (climate factor). Rosenthal (1974) stated it this way:

> The factor, then, can be viewed as very much related to how much active teaching occurs but specifically omitting the variable of how much new material is presented. . . . Direct criticism therefore does not conflict with the operation of the climate factor. (p. 18)

As stated earlier in Beez' (1970) experiment, teachers gave only four reinforcements to the children of whom they expected less in comparison to seven reinforcements to children of whom they expected more.

Lanzetta and Hannah (1969) found that teachers may teach more clearly to those students with high expectations. Psychology students were asked to teach a task to their peers, some whom the teacher expected to show high potential, and some whom the teacher expected to show a low potential for learning. All subjects were told of the teachers' expectations. For every task response, the teacher had to respond with five feedback choices: a strong or mild electric shock, a neutral light, and a large or small monetary award. Each subject gave the same number of correct

and incorrect responses. Results showed that when the learner (fast or slow) gave the exact number and correct number of responses, the high expectation learner still received a high level of positive reinforcement. In addition, when the high potential learner gave an incorrect answer, the subject received a strong level of shock compared to the low potential learner, indicating that when teachers expect more, they send clear signals to let the student know that they expect more.

In other experiments by Rubovitz and Maehr (1971), Rothbart et al. (1971), and Kester (1971) children were given more attention time than the control children. There was no difference in the amount of positive or negative reinforcement (quality of interactions). but there was a difference in the amount of attention shown to children of both groups.

Results of the studies by Brophy and Good (1970), Meichenbaum et al. (1969), and two studies by Rubovitz and Maehr (1971, 1972) did not find any effect of teacher expectations on time or attention. Rosenthal (1974) suggested that:

> At least under some conditions, differential teacher attention may serve to mediate teacher expectations. If further research does not provide additional support for the attention hypothesis it will still be necessary to separate out those components of 'attention' that are not part of the Climate or Feedback Factors. (p. 20)

Input

Teachers tend to teach more to children of whom they have higher expectations than to children of whom they have lower expectations (Rosenthal, 1974, p. 20). Five studies compiled

by Rosenthal support the hypothesis of the Input Factor. Four of these studies (Beez, 1970; Brown, 1969; Carter, 1969; McLean, 1970) followed, more or less, the same paradigm. Series of words, paired and associated learning tasks, were used to teach pupils in expectancy and control groups. The results showed that when children were labelled as brighter or expecting to learn more, they were taught or exposed to more words or associations compared to students who were believed to be dull. In the fifth study, Rist (1970) concluded that the teachers, on the basis of physical looks and educational information, placed children in groups of low promise or high promise and the brighter group (over a three year period) was taught more than the dull group. Rosenthal (1974) stated that these five experimental results:

> . . . gives us considerable confidence in our conclusions that one factor in the mediation of teacher expectations may well be how much the teacher teaches to those of whom she expects little. (p. 22)

Output

Teachers tend to encourage greater responsiveness from students of whom they expect more. Rosenthal (1974) elaborates on this factor as follows:

> Such encouragement might take form of calling more often on those children of whom more is expected, asking them harder questions, giving them more time to respond, and prompting and shaping partially correct responses so that they become more correct. (p. 22)

Rosenthal has stated that the four factors are distinguishable from one another but not necessarily independent from one another. However, this last factor is more aligned with the

Feedback Factor. If children respond more they may also get more feedback on the correct or desired response.

Davis and Levine (1970) discovered that when teachers were told of children who showed unusual intellectual development, they called on those children 60% more often than those children who were in the control group. Audio-visual tape recordings were analyzed as to the number of times teachers asked children questions. Children with high expectations were called on at an average of 11.9 times, whereas the children with no special expectation were called upon on an average of 7.3 times.

Gess (1969) followed a similar pattern but divided his subjects into three levels of high, average, and low groups based on teachers' opinions and expectations. Teachers were videotaped and the number of questions were coded showing that there was little difference between the two lower groups. The results also showed that teachers called upon the high-expectancy group 50% more than the two lower groups.

In an industrial setting, King (1971) observed industrial supervisors' expectations and found that workers were given more demanding assignments if more was expected of them compared to workers in the control group. Also, workers of the experimental group were watched and supervised more closely than the control group.

Rowe (1969), supporting the mediation of the Output Factor, was interested in how long teachers waited for responses from students before repeating the question, asking another student

to answer the question, or asking the first student another question. Rowe found that teachers waited longer when better students were asked questions. Rowe also found that when teachers were told this fact, that they then purposely waited longer for the slower students to respond. As a result of waiting, the slower students increased their responses. Children were given more opportunity to respond or to show their knowledge if the teacher expected more of them.

Rosenthal (1974) reviewed 13 studies that reported the Output Factor and only one study gave results in the opposite direction. In summarizing the Four Factor Theory, Rosenthal (1974) stated:

> When viewed as dependent variables arising from differences in teachers' expectations, these factors can be measured and correlated in future studies. When viewed as independent variables leading to differences in pupil performance, these factors can still be measured and correlated but they can be also varied experimentally, independently of one another. One may hope that in the not too distant future we may know what sources and kinds of teacher expectations lead to the operation of these and other factors and what effects these factors have on pupil performance. (p. 24)

E. Objections and Critical Reviews of Rosenthal's Work

The publication, Pygmalion in the Classroom, not only encouraged researchers to replicate the study, but the report also created objections from psychologists and educators on a number of grounds (Thorndike, 1968, 1969; Jensen, 1969; Elashoff & Snow, 1971). Since the book ". . . addressed a major social problem and received nationwide attention" (Elashoff & Snow,

1971, p. vi), a number of reviews were written both supporting and objecting to the study and the publication.

In 1966, at the American Psychological Association Symposium, Rosenthal and Jacobson gave their report on Pygmalion. Gage, serving as a discussant, indicated his skepticism by citing weaknesses in the design, and weaknesses and discrepancies in the measurement and analysis of the experiment. In 1967, Gage was asked to review the manuscript, and as in the symposium in 1966, criticized the report severely. He had stated that the book received high praise from almost all reviewers. But, Gage (1971) also said that ". . . most of the reviewers were untrained in psychological measurement and statistical analysis" (Foreword). In addition, Thorndike (1969), Jensen (1969), Snow (1969), and Elashoff and Snow (1971) questioned the validity of the study's data and conclusions and as a result of these criticisms, Elashoff and Snow (1971) collaborated on a publication to point out the questionable nature of the Rosenthal and Jacobson study. Gage (1971), grateful for this publication, stated:

> Now that the Rosenthal-Jacobson work has been thrown in doubt, one can only hope that the whole business will not . . . undermine confidence in psychological research. (Foreword)

Elashoff and Snow (1971) published Pygmalion Reconsidered, a case study of Pygmalion in the Classroom (Rosenthal and Jacobson, 1968). The authors (1971) stated that they chose this study for detailed examination for two reasons:

> First, it addresses a major social problem, has received nationwide attention, and has prompted a number of similar studies in the area; second, its

> basic design, measurement problems, and the statistical procedures used in its analysis and re-analysis are typical of those encountered frequently in educational behavioral science. (p. vi)

Elashoff and Snow criticized the Pygmalion study as a research report suggesting that the report as a whole was inadequate. They described the design, basic data, and the analysis as incomplete. They further stated that there were inconsistencies between tests and tables, dramatic conclusions, inaccurate or incorrect statistical discussions, and misleading analyses, all contributing to a generally misleading impression of the study's results (Elashoff & Snow, 1971, p. 6).

The authors were concerned about the amount of publicity and attention that the book had received because they believed the report did not contain a complete understanding of the data and results. They were especially concerned about the interpretations and conclusions, design and sampling problems, and measurement problems stating that ". . . tests and tables are inconsistent, conclusions are overdramatized, and variables are given prejudical labels" (Elashoff & Snow, 1971, p. 10).

Elashoff and Snow's argument concerning the design and sampling plan was that the sampling was ill defined, the procedure for assignment to treatment groups was obscure, and an imbalance was deliberately created. Rosenthal selected at random 20% of the children in his study to be the experimental group, and the number of experimental children in each classroom ranged from one to nine. Elashoff and Snow (1971) indicated that such a ". . . lack of equality in the number of experimental

children per classroom means that some classes have too few experimental children to make analysis within classrooms feasible" (p. 21). Because of subject loss and inequalities of the group initially, the authors (1971) felt that the experimental group and the control group could not be regarded as ". . . representing comparable groups" (p. 23).

In the measurement section of their analysis, the authors thought that the use of TOGA alone was inadequate to measure intellectual growth and questioned the reliability and validity of Rosenthal's criterion measure. Rosenthal did not attempt to relate the TOGA scores to other acknowledged intellectual measures. Elashoff and Snow (1971) stated:

> It is not clear how valid the TOGA IQ measures themselves are as a measure of intelligence or achievement or how valid changes in TOGA IQ scores are as a measure of intellectual growth. (p. 39)

In the discussion of "Interpretations and Conclusions," the authors (1971) were concerned about labelling dependent variables such as "intellectual growth" and "expectancy advantage." These labels, they said ". . . presume too much and make interpretations before any effects are found, as well as imply differences are always positive" (p. 10). Elashoff and Snow (1971) also indicated that there was a ". . . clear tendency to over-generalize the findings" (p. 11). They implied that whenever Rosenthal had contradictory results, the conclusions sounded quite different.

The authors offered to their readers a number of recommendations for further research on teacher expectancy effects, because

according to Elashoff and Snow (1971), Rosenthal and Jacobson's study was ". . . inadequate in the choice of analytic procedure, in the choice of criterion measures, and in the attention paid to basic data" (p. 43).

As a result of Elashoff and Snow's critical analysis, Rosenthal and Rubin (1971) retaliated with a point by point rebuttal to Elashoff and Snow's publication, indicating that the criticisms were unsound. Rosenthal and Rubin (1971) concluded that Elashoff and Snow's re-analysis only confirmed the Pygmalion study and stated that:

> Although there were among the ES criticisms a few useful notions which we employed in this reply, in the main, the numerous criticisms advanced in ES were neither sound nor constructive. (p. 155)

In addition to Elashoff and Snow's publication, others in the field of education and psychology were compelled to review Pygmalion in the Classroom. The most scathing attack of the Pygmalion report was by Thorndike (1968) who stated in his opening paragraph:

> In spite of anything I say, I am sure it will be a classic--widely referred to and rarely examined critically. Alas, it is so defective technically that one can only regret that it ever got beyond the eyes of the original investigators. Though the volume may be an effective addition to educational propagandizing, it does nothing to raise the standards of educational research. (p. 708)

Thorndike's objection, like Elashoff and Snow's, was not the efficacy of the self-fulfilling prophecy nor the other previous research on the expectancy effect. Thorndike's (1968) main objection and the main point of his review was directed at the

". . . adequacy of procedures (of data gathering and data analysis), and the appropriateness of the conclusion drawn from the study" (p. 708).

In another article, in the Harvard Educational Review, entitled, "How Much Can We Boost IQ and Scholastic Achievement?," A. R. Jensen (1969) criticized the statistical procedures that Rosenthal and Jacobson used. Jensen stated that the administration of the IQ tests was unreliable and, therefore, he questioned the reliability and validity of the results.

Not all criticisms were directed toward the statistical design of the Rosenthal and Jacobson study. Other writers (Aiken, 1969; Coles, 1969; Kohl, 1968) expressed their views on the significant findings which would have implications for further studies and research.

Aiken (1969) reviewed Rosenthal's Pygmalion report in the book review section of Education and Psychological Measurement. In his review (unlike Thorndike's, Jensen's, and Elashoff & Snow's), he did not delve into the experimental design per se, but only reported Rosenthal's findings. However, Aiken (1969) did mention the fact that ". . . the control groups were much larger than the experimental groups, that gain scores were open to question, and that many of the significant differences may have been caused by the scores of only a few children" (p. 228). He stated that because of the above discrepancies there would be many critics attacking the study, but he also emphasized the point that in no way could the fact be denied that there had

been significant findings that could not be dismissed or destroyed. Aiken (1969) believed that the results of Rosenthal's study pointed to ". . . a need for a reassessment and more careful analysis of the effects of teacher behavior, both verbal and nonverbal, and teacher attitude on the attitudes, self-concept, and performance of school children" (p. 228)

An article by Robert Coles (1969, 1971), in the New Yorker Magazine, was reprinted in Elashoff and Snow's book. As in Aiken's review, the author did not question the way in which the study was designed but commented only on the far-reaching implications of such a study. Coles (1971) was very much impressed by the results and suggested further research ". . . to discover how teachers go about letting children know they have a special destiny" (p. 80). He believed that nonverbal nuances and signals (look, touch, facial expression), as well as verbal communication transmitted by the teacher, set the stage for messages received by the children, who, in turn, satisfied those teachers' messages.

In the New York Review of Books, Herbert Kohl (1968) also discussed the important implications of the Rosenthal study and related especially to the subject of tracking and grouping abilities. He agreed that the self-fulfilling prophecy worked especially in the ghetto schools and in schools where there were tracking systems (ability tracks) and said:

> Almost without exception, the grouping according to track is self perpetuating; and the students usually remain in the same track throughout their school career. (p. 31)

Although Kohl agreed that Rosenthal's work was gratifying, he did not condone the research methods used by Rosenthal. He objected to the manipulation of teachers (whose cooperation was enlisted in bad faith), to the neutral and cold observation of manipulation of human beings, and to the underhanded tactics that were used to get information. Kohl (1968) concluded by saying:

> This study does not reveal what teachers who have been studied feel nor whether they have learned something about themselves that could have some effect . . . . Surely there must be a more direct way of confronting teachers with their attitudes, and studying them in a more direct way. (p. 31)

In April of 1969, Snow (1969), in his article in Contemporary Psychology, criticized the Rosenthal-Jacobson study by saying that their research would have been ". . . judged unacceptable if submitted to an APA journal" (p. 197). In rebuttal to Snow's statement, Rosenthal (1970) reported to the Journal of Contemporary Psychology with an article entitled, "Another View of Pygmalion" (p. 524). He offered some evidence to Snow of the acceptance of the Pygmalion study. Rosenthal stated that his study received first prize of the 1967 Cattell Fund Award presented by Division 13 of the American Psychological Association. He also stated that the editors of the book, Social Class, Race and Psychological Development (sponsored by Division Nine of the American Psychological Association) asked Rosenthal to submit his Pygmalion research for inclusion in the volume. (One of the authors, A. R. Jensen, objected to the inclusion of the article.)

Rosenthal has taken note of the objections and criticisms aimed at his study, Pygmalion in the Classroom. But he also has stated that there is a considerable amount of evidence to show that ". . . one person's expectation of another person's behavior can come to serve as a self-fulfilling prophecy" (Rosenthal, 1974, p. 24). He has now suggested that as these studies accumulate, they become less useful since all they do is further the high probability that such effects do occur. Elashoff and Snow (1971) emphasized the same point when they said, ". . . the quest for further research is not whether there are expectancy effects but how they operate in school situations" (p. 64).

Summary

Rosenthal optimistically has stated that enough research (his 30-study review) has accumulated on the mechanisms that mediate interpersonal expectation effects to make use of his Four Factor Theory. In summarizing his work, Rosenthal (1974) stated:

> It will not do to conduct the experiments that will answer all the questions. It will not do even to conduct two, three, or four experiments in hopes of finding the answers. The behavioral sciences, particularly when operating with molar, social interactional variables, do not work that way. Many studies are needed, conducted in different settings, both laboratory and field, by different workers, in different centers of research. (p. 24)

CHAPTER III

PROCEDURES

The purpose of this study was to identify specific and differential teacher behaviors that affect student behavior based on Rosenthal's Four Factor Theory (1974, pp. 14-24). Specific purposes included the following:

1. The identification of teachers' differential treatment of students according to the teachers' expectations (high or low).

2. The identification of students' different responses according to the teachers' expectations (high or low).

3. An investigation of the factor, Touch, as another possible identifying mechanism responsible for the Pygmalion Effect.

This chapter presents the procedures used in obtaining data pertinent to teacher expectations. The procedures for this study involved preliminary preparation, collection of data, and the intermediate stages (separate tabulations of some categories) of data preparation. The procedures are examined in three sections. The preliminary preparation section includes the development of an adaptation of Brophy and Good's (1969) Manual for Coding Classroom Behavior, the selection and training of judges, and the statistical method used to obtain acceptable reliability and intercoder agreement. The second section describes the collection of data which includes selection of school, teachers, classes, and

subjects, observational procedures, and testing dates. In the third section, there is an explanation of the separate tabulations and development of summary sheets used for the preparation of data for analysis.

Preliminary Preparation

Selection of a System for Observing Teacher-Student Behavior

The Brophy and Good (1969) Teacher-Child Dyadic Interaction System is a system which studies dyadic interactions between teachers and students in classrooms. Interactions between the individual student and the teacher are recorded and analyzed separately so that the student rather than the class is treated as the unit of analysis. The system does not involve coding everything that happens in the classroom such as interactions with groups of students or interactions with the class as a whole unit. The emphasis is on the word dyadic; the interaction of the teacher with a single student.

The Brophy and Good (1969) interaction analysis system was developed for the specific purpose of studying interclass individual differences and differential performance expectations by teachers. The authors felt that none of the other coding systems devised for studying classroom behavior using the classroom as a unit of analysis adequately reflected a teacher's specific behavior toward a particular individual or subgroup. The authors believed that some teacher variables such as teacher warmth and teacher indirectedness were not being evaluated

accurately. For example, if the variable, teacher warmth, was used as a measure of teacher effectiveness, and if the classroom was used as the basic unit of analysis, the measure would inaccurately portray the teacher's general behavior as well as the degree of warmth shown toward individuals. A teacher could score high on teacher warmth from an observation system using the class as the unit of measurement, but the investigator would not be able to report to whom this warmth was directed.

The investigator used and adapted Brophy and Good's (1969) observation system for two reasons: (a) Dr. Robert Rosenthal suggested that this system would be appropriate to use for the present study dealing with the communication of teacher expectations, and (b) the authors (Brophy & Good, 1969) stated that ". . . the basic research methodology for coding dyadic teacher-child interaction in this manual can be extended to the study of almost any kind of behavior" (p. 69). Before a final commitment was made to use the Brophy and Good system, the investigator practiced the system in physical education classes to ascertain any changes that needed to be made in the system or on the coding sheet.

Adaptation of the system. Brophy and Good (1969) present five different types of dyadic interaction situations in their observation system. They are: Response Opportunities, Recitation, Procedural Contacts, Work-Related Contacts, and Behavior Contacts. For the purposes of the present study, some categories were added. Two of the five major divisions were deleted from the system. The Recitation and Work-Related Contacts were not considered because

the investigator decided that any behavior appropriate to these two particular situations could be absorbed into the Response Opportunities section. For example, demonstration of a skill was considered comparable to the Recitation situation in some instances (when the teacher gave skill tests). The investigator was informed that no skill tests would be given during the two-week observation period. On the basis of this information, the investigator deleted the section. The following pages describe the five major dyadic interactions. Explanations for the adaptations and deletions follow each of the major divisions. A more complete description of both the categories and subcategories in each division can be found in the Appendix.

1. Response Opportunities. The student publicly attempts to answer the question posed by the teacher. The teacher affords this opportunity to the student and deliberately encourages the student to respond. The nature of the interaction involves a question (teacher), followed by an answer (student), followed by appropriate feedback (teacher).

Adaptations. General Task was added to the system for ease of coding and to minimize any confusion relating to the type and level of question. The student publicly attempts to respond to a movement task demanded by the teacher. The movement task demanded by the teacher is comparable to a verbal question and the movement task executed by the student is comparable to a verbal answer. The subcategory, Praise, was changed to Climate. The subcateory, Praise, was considered comparable to Climate

and only a word change was necessary. The subcategory, Rephrase-Clue, was extended to include Suggestion and Correction. A new subcategory, Student-Initiated Questions, was included to take into account any questions posed by the students. In the Quality of Answer-Movement Category, the term Don't Know was added so that the coders could distinguish between the student who responded with the answer, Don't Know, and the student who did not answer or acknowledge the teacher's question. General Touch, Incidental Touch and Assist Touch were added to the appropriate categories in the Response Opportunities section.

2. Recitation. The student reads aloud, describes some experience, or makes an oral presentation.

Deletion. The Recitation section was not needed and any description or short presentation that the student was required to make was considered a response opportunity and entered in the Response Opportunities section in the appropriate categories.

3. Procedural Contacts. Procedural Contacts are Student-Created or Teacher-Afforded Contacts. The student receives permission to move to another place, to move, put away, or give out equipment, or the teacher asks the student to perform such duties.

Adaptation. Procedural Touch was added to both the Student-Created and the Teacher-Afforded sections.

4. Work-Related Contacts. An interaction which involves homework, seatwork, or other written work done by the student.

Deletion. This type of behavior was considered part of a response opportunity. If the teacher told the students to work

on a particular movement task and then assisted them, the movement task was coded in the Response Opportunities section and considered a public interaction between the student and the teacher.

5. Behavioral Contacts. The teacher disciplines the student or makes specific comments to the student concerning his or her behavior.

Adaptation. Behavioral Touch was added to the Behavioral Contacts section.

Process of Coding

Every interaction between the teacher and the individual is coded. The sequential nature of the teacher-student interaction is also kept intact so that cycles of initiation and reaction are not lost in the coding process. One of the most important features of the system is knowing who created the interaction; the teacher or the student. The most difficult part of the coding system is the Response Opportunities section where there may be as few as four and as many as seven or more checks for one interaction. If the teacher offers sustaining feedback to the student, an infinite number of checks may be entered on the coding sheet for one original response opportunity showing a sequential order of coding events. The coding of sequential events gives information concerning the type of teacher feedback focusing on the quality of contact. Both the quantity of contact (the number of different kinds of interactions) and the quality of

contact (what the teacher says or does during these interactions) can be studied separately.

Response opportunities. Each response opportunity which is coded requires five separate entries or checks: the student's identity number, the type of response opportunity, the level of question asked, the quality of the student's answer, and the type of feedback response from the teacher. The last item, teacher feedback, is more complex because there may be more than one feedback response during an interaction.

Basically, the sequence of events is built into the coding sheet. The coder starts from the left side of the sheet and then moves to the right, coding the decisions that take place in the order that they occur. Below is an example of the sequence of events that are coded when the teacher and an individual student interact.

First, the coder indicates the identity of the student by entering the student's number in the Type of Question column (Discipline, Direct, Open). This entry both identifies the student and the type of question. The next check is entered in the Level of Question column (Process, Product, Choice). The next entry indicates the quality of the answer (Correct, Incorrect, No Response, Don't Know), followed by a check in the appropriate feedback column. In both the Terminal and Sustained Feedback categories, there can be more than one codable behavior during an interaction. The sequence and the nature of the feedback must be noted in order. For example, the first entry would

be indicated by a check or a number one. If the teacher sustains the feedback, the entry is noted with a number two, followed by number three, and so on, until the teacher terminates the response opportunity. The nature of the response opportunity is indicated by a check in the row below showing the close of the original response opportunity and the beginning of the follow-up response opportunity. On the row below, the entries would indicate the level of question, the quality of the answer and the type of feedback. The student's identification number would not be entered again in the Type of Question column since identification of the student is coded only for an original response opportunity and not for a follow-up response opportunity. Some examples of both Terminal and Sustained Feedback are presented in Tables 1 and 2 for further clarification.

Table 1

Coding Example in Response Opportunities for Terminal Feedback

Student:	Identification Number 21
Question:	Tommie, does the term "bogey" mean one over par or one under par?
Answer:	One over par.
Feedback:	That is correct. Very good!
*Code:	Number 21 is entered in the Direct column, and subsequent checks are entered in the Choice, Correct (+), Affirms Right, and Climate (++) columns.

*See sample coding sheet on page 62.

Table 2

Coding Example of Response Opportunities for Sustained Feedback

Student:	Identification Number 24
Question:	Millie, does the term "bogey" mean one over par or one under par?
Answer:	One over par.
Feedback:	That is correct. Very good! Do you know what the term "one under par" means?
Answer:	Yes! It's a birdie.
Feedback:	Yes, Millie. Very good!
*Code:	Number 24 is entered in the Direct column, and subsequent checks are entered in the Choice, Correct (+), Affirms Right, and Climate (++) columns. Continuing with sustained feedback, another check is entered on the same row in the New Information column. On the next row below, checks are entered in the Product, Correct (+), Affirms Right, and Climate columns.

*See sample coding sheet on page 62.

Note: The identification number of the student would not be entered on the row below since this is a follow-up response opportunity.

Procedural contacts. Procedural dyadic contacts are coded separately according to whether they are student-created or teacher-afforded. For the afforded interactions, a number identifying the student is entered in the Feedback column. If the teacher touches the student during this interaction, a check is entered in the Touch column. For Student-Created, Procedural Contacts, the student's identification number is entered in either the Praise (++), Feedback, or Criticism (-) columns. The Feedback column is

checked when the teacher gives feedback to the student without either praise or criticism. See Table 3 below.

Table 3

Coding Example of Teacher-Afforded and Student-Created Procedural Contacts

	Teacher-Afforded
Student:	Identification Number 18
Teacher:	Joy, please help Sally move that bench.
Code:	Identification number is entered in the Feedback column
	Student-Created
Student:	Miss Smith, would you like me to give out the equipment?
Teacher:	Yes, Joy. You are thoughtful to volunteer. (Teacher touches Joy on the arm.)
*Code:	Identification number is entered in the Praise column and a check is entered in the Touch column.

*See sample coding sheet on page 62.

Behavioral contacts. Behavioral contacts are coded whenever the teacher comments on a student's behavior. If a student is misbehaving, the identification number is entered in either the Warning or Criticism columns. If the teacher makes a positive statement relating to a student's behavior, the identification number of the student is entered in the Praise column. If a teacher touches a student when giving behavioral feedback, a check is entered in the Touch column along with the accompanying feedback. A clarifying example is presented in Table 4.

Table 4

Coding Example of Behavioral Contacts

Student:	Identification Number 17
Teacher:	Betsy, I have told you before to be quiet!
*Code:	Identification number is entered in the Warning column.

*See sample coding sheet on page 62.

General coding conventions. The authors (Brophy & Good, 1969) have stated that ". . . since the system involves objective coding of observable behavior, its validity is insured automatically if it is reliably applied according to the instructions in the manual" (p. 104). Coding rules were established by the authors to ensure ". . . the validity of data in studies of teacher communication of expectations through differential behavior toward different students" (p. 48). The following conventions were applied by the investigator and judges:

1. Nothing was coded if the judges could not identify the student interacting with the teacher. The judges were instructed not to guess the identity of a student. This rule minimized any contamination of observation data by the expectations of the coder.

2. The teacher's intent was the most important consideration for the determination of proper coding when more than one category could apply. For example, if the teacher considered a correct answer to be wrong, the judges coded what the teacher intended.

1. Class ________ 2. Date ________ 3. Start _____ 4. _____ 5. Activity __________ 6. Attendance _________

7. Observer ________________________________ 8. Page ____________ of __________.

Response Opportunities																																						
Type of Question					Level of Question					Ans-Movt				Atmos		Terminal Feedback										Systm			S-Prc				Tprc		Behavior			
DISC	DIRECT	OPEN	CALL	STUD INIT	GEN TASK	PROCESS	PRODUCT	CHOICE	SELF REF	+ CORRECT	- WRONG	NO RESP	DONT KNOW	GEN TOUCH	INC TOUCH	++ CLIMATE	AFF RIGHT	NO FDBCK	NEG WRONG	-- CRIT	ASST TOUCH	PROCESS	GIVE ANS	ASK OTHER	CALL	REPEAT	REPH SUGG	NEW INFO	PROC TOUCH	+	FB	-	TOUCH	FB	+	W	-	TOUCH
	21							✓		✓						✓	✓																					
	24							✓		✓						✓	✓											✓										
							✓			✓						✓	✓																					
																																	1	18				
																													✓	18								
																																				17		

Expectations:

Remarks:

Figure 1

Coding Sheet

The observers were instructed not to determine what was right or wrong, but what the teacher indicated in her feedback responses.

3. Coders were instructed to study carefully all ambiguous situations. All borderline situations were thoroughly discussed and learned. In all categories, the general procedure for resolving borderline or ambiguous situations was to code the category which implied less about the communication of teacher expectations. For example, if there was an indecision concerning a direct question and a discipline question, the direct question would be coded as the type of response opportunity. The direct question implies less about the teacher's intent than the discipline question.

4. All teacher feedback reactions were coded in the sequential order in which they occurred. For example, if the teacher said to a student: Yes, Mary, that is good!, the coder would check or use the number one in the Affirms Right Column, then place a number two in the Climate Column.

5. All dyadic interactions (procedural and behavioral) were coded as single interactions regardless of the length of the interaction between the student and the teacher.

6. All unforeseen types of responses were noted by the judges and discussed with the investigator after each class.

7. Judges were instructed to be especially careful about double-coding behavior. For example, praise and criticism categories are in more than one division. Therefore, if a teacher initially criticized a student in a procedural contact and then

made some comment upon his behavior, the judge would check only the column in the procedural contact and not the behavioral section.

8. Coders were instructed not to repeat the student's identification number during sustaining feedback. The only way to get an accurate count of original response opportunities was to count the number of times the identification number was entered in the response opportunity section.

9. Coders were instructed that all response opportunities were ended in the Terminal Feedback Column. If the teacher ignored a student and gave no feedback after a question, the coder entered a check in the No Feedback Column.

Most of the conventions and instructions suggested by Brophy and Good (1969) were followed by the investigator. Additional rules were necessary for the present study because of some of the adaptations and the addition of categories into Brophy and Good's system. A detailed description of Brophy and Good's (1969) interaction analysis system can be obtained from The Research and Development Center for Teacher Education at the University of Texas at Austin.

Selection and Training of Judges

Three full-time graduate students from the University of North Carolina at Greensboro were asked to be observers for this study. No criteria were used for the selection of the observers. Two of the students were doctoral candidates and one student was a master's candidate; none of them had previous training in

observational systems. These three students were recommended by another graduate student because they were either not teaching classes or did not have a heavy course load and thus could commit a block of time to the investigator.

A preliminary meeting was held by the investigator to explain the purpose of the study and to describe the procedures and practices necessary for training. Prior to the meeting, a packet was given to each of the graduate students. The packet included a detailed description of the observational system that was to be used for the study and a letter asking them to study the system carefully so that they could be ready for a question and answer period at the first meeting. The selectees indicated that they would be able to learn the system, but their main question was the time involved for training. The investigator could only rely on the information from Brophy and Good's (1969) manual which indicated a two-week training period. For this study, the training period lasted for six weeks. This was due to the complexity of the system coupled with the nature of the physical education activities.

A training schedule was made according to the observers' available time. Most meetings were scheduled from 9 a.m. to 12 noon. Unless there were extenuating circumstances, training sessions were scheduled for once a week for the first four weeks. During the latter part of the training session (after four weeks), meetings were increased to two times per week. Each session lasted for at least two, and no more than, three hours. The total

training time amounted to 29½ hours. The observational sessions (testing days) involved 24 hours (4 hours/day for 6 days). Total training time and observational time approximated 53½ hours. The training schedule which includes the time, date, and place is shown in Table 5.

Videotape recordings were taken by the investigator at the University of North Carolina at Greensboro and used for the six-week training period. Three different physical education classes were filmed: a fencing class, a social dance class, and a body mechanics class. These three classes were selected because of the uniqueness of each class and the obvious differences between the activities. The length of the videotapes (classes) was approximately 40 minutes. In addition to the videotapes, verbal sessions were held during which sample questions (over 100) were given to the judges for study. Observations of live classes at the University were also coded. In addition, the investigator and judges went to the Allen Junior High School to observe a class to get a coding experience similar to that which would be involved in the study. A videotape recording was made while the judges were observing the class. The judges met together to observe additional college classes to gain more experience.

For all training sessions, the judges and investigator met in the videotape recording room. Each session started with some verbal discussion of the categories and clarifications of particular disagreements. The judges then looked at a videotape for a short period of time. The investigator would stop the film

Table 5

Training Schedule for Judges

Date	Place	Time
September 27, 1976	Student Lounge	9:30-11:30 a.m.
October 4, 1976	Videotape Recording Room	9:30-11:30 a.m.
October 7, 1976	Videotape Recording Room	9:00-12:00 noon
October 14, 1976	Videotape Recording Room	9:00-12:00 noon
October 20, 1976	Videotape Recording Room	9:30-12:00 noon
October 21, 1976	Videotape Recording Room	10:00-12:00 noon
October 27, 1976	Videotape Recording Room	10:00-12:00 noon
October 28, 1976	Videotape Recording Room	10:00-12:00 noon
October 29, 1976	Videotape Recording Room Pre-test on one film	9:00-12:00 noon
November 1, 1976	Rosenthal Gymnasium Dr. Barrett's Gymnastic Class	10:15-11:15 a.m.
November 4, 1976	Videotape Recording Room Allen Junior High School Videotape Recording Room	10:00-11:00 a.m. 11:15-12:30 p.m. 12:45-2:30 p.m.
November 7, 1976	Videotape Recording Room	9:00-11:30 a.m.

either when asked by the judges or when the investigator felt that there needed to be some explanation of a particular interaction. The investigator would then discuss any points of confusion and clear up any obvious disagreements among or between the judges. After clarification of disagreements, the same portion of the tape would be shown to see if the judges could discern the correct behavior responses and come to an agreement.

The uniqueness of the observational system is that all interactions between the student and the teacher are coded. Therefore, it was especially difficult for the judges (in the training sessions) to code the interactions of a 30 or 40-minute tape. In the actual testing sessions, the judges were responsible for only eight students each so that only a percentage of interactions of the total class was recorded by one judge. For the training sessions, it was necessary to code all interactions that were on the tape. The students did not have numbers, nor did they wear identifying colors in any of the training films taken at the University. When the judges became proficient at coding for a specific period of time, the time of observation was increased until they could observe at least 20 minutes of activity on the tape.

A special coding sheet was adapted from the Brophy and Good (1969) manual by the investigator for the training session. This sheet was also used for the final testing session and data collecting session. (See Appendix)

After four weeks of training, the judges were given a pretest on the fencing tape. The investigator was interested in

assessing the data to get some indication of intercoder agreement. Brophy and Good (1969) have recommended that 80 percent intercoder agreement be attained before coders begin to work alone. The results of the fencing tape indicated a satisfactory intercoder agreement. One week later, the judges observed an actual college class. Another assessment was taken and in most categories, the 80 percent agreement figure was not reached. The judges discussed their points of confusion and suggested more actual class training. During the sixth week, the judges observed a junior high school physical education activity class. A film was taken of the class at the same time that the judges were coding. It was the intention of the investigator to assess the objectivity of the judges using the class data and the film data. However, the number of coded behaviors in the class did not coincide with the number of coded behaviors on the tape. The judges stated that they were sure that they had recorded some behaviors in the actual class that could not be readily seen on the tape. The investigator made the decision to use the film for intercoder agreement. Data on the junior high physical education activity tape were assessed on Thursday of the sixth week for the final testing session. The judges were tested again on the same tape three days later. The data were treated statistically to estimate satisfactory coding agreement of 80 percent.

Statistical Method Used for Assessing Intercoder Agreement

The coded observations were treated statistically to determine the reliability and objectivity of the three judges. Brophy and Good (1969) recommended using the percentage of agreement score to determine intercoder agreement. Percent agreement is determined by the ratio of exact agreement between coders to the combined total of exact agreements, plus omissions (one coder coded and the other did not), plus disagreements (both coders coded but disagreed on the coding).

Reliability. Reliability refers to the consistency of measurement or the ability to measure the same thing on two different occasions (Safrit, 1973). Each of the three judges coded the same tape on two different occasions three days apart. Percent of agreement was obtained using the formula suggested by Brophy and Good (1969). Intercoder agreement of 80 percent recommended by Brophy and Good (1969) was accepted for this study. In three of the categories, No Feedback, Rephrase, and New Information, the percent of agreement was below the recommended 80 percent. This was due to the small number of interactions. In the three categories, there were no more than four interactions. Because of the training and knowledge of the three judges, the investigator accepted the low scores and did not consider these scores indicative of the judges' training, but rather due to the small number of behavior responses. When no behaviors were recorded, this indicated 100 percent agreement and was shown as ** (Lunt, 1974). The percentage of intrajudge agreement is shown in Table 6.

Table 6

Reliability Index: Percentage of Intrajudge Agreement in Each Category of Response Opportunities

Category	Sub-Category	Judge A	Judge B	Judge C
Student	Discipline	**	**	**
	Direct	100	100	100
	Open	100	100	100
	Call	100	100	100
	Student-Initiated	100	100	100
Level of Question	General Task	100	100	100
	Process	**	**	**
	Product	93	95	94
	Choice	**	**	**
	Self-Reference	**	**	**
Answer-Movement	+ = Correct	100	94	93
	- = Incorrect	100	100	100
	No Response	**	**	**
	Don't Know	**	**	**
Atmosphere	General Touch	**	**	**
	Incidental Touch	**	**	**
Terminal Feedback	++ = Climate	100	100	100
	Affirm Right	93	86	93
	0 = No Feedback	67	75	100
	Negate Wrong	100	100	100
	$\underline{-}$ = Criticism	**	**	**
	Assist Touch	**	**	**
	Process	**	**	**
	Give Answer	100	100	100
	Ask Other	**	**	**
	Call	**	**	**
Sustained Feedback	Repeat	**	**	**
	Rephrase-Suggest	100	100	75
	New Information	100	100	75
Created Procedure Student	Procedure Touch	**	**	**
	+	**	**	**
	-	**	**	**
	Feedback	**	**	**

Table 6 (Continued)

Category	Sub-Category	Judge A	Judge B	Judge C
Afforded Procedure Teacher	Procedure Touch	**	**	**
	Feedback	100	100	100
Behavior	+	**	**	**
	-	**	**	**
	Warning	**	**	**
	Behavior	**	**	**

** = no recordings made; considered 100 percent agreement.

Objectivity. Interjudge agreement or objectivity refers to the ability of different judges to measure the same behavior responses with consistency. The judges' scores from the same tape were tested to determine agreement between judges. Each judge was paired with the other two judges to ascertain interjudge agreement. The same index that was used to obtain the reliability score was also used to determine the intercoder agreement. Again the 80 percent agreement recommended by Brophy and Good (1969) was used as the standard. Each judge agreed with every other judge and exceeded 80 percent in all categories except for the three categories previously mentioned for intrajudge agreement. The low percentage score was due to the small number of behavior responses and the discrepancies were omissions rather than disagreements. The percentage of interjudge agreement is shown in Table 7. These scores were accepted by the investigator. The asterisk symbol ** indicated that no recordings were made and

Table 7

Percentage of Interjudge Agreement in Each Category of Response Opportunities

Category	Sub-Category	Judge AB	Judge AC	Judge BC
Student	Discipline	**	**	**
	Direct	100	100	100
	Open	100	100	100
	Call	100	00	00
	Student-Initiated	100	100	100
Level of Question	General Task	86	86	100
	Process	**	**	**
	Product	83	94	89
	Choice	**	**	**
	Self-Reference	**	**	**
Answer-Movement	+ = Correct	94	94	88
	- = Incorrect	100	100	100
	No Response	**	**	**
	Don't Know	**	**	**
Atmosphere	General Touch	**	**	**
	Incidental Touch	**	**	**
Terminal Feedback	++ = Climate	75	100	75
	Affirm Right	100	100	100
	0 = No Feedback	67	33	50
	Negate Wrong	100	100	100
	-- = Criticism	**	**	**
	Assist Touch	**	**	**
	Process	**	**	**
	Give Answer	100	100	100
	Ask Other	**	**	**
	Call	**	**	**
Sustained Feedback	Repeat	**	**	**
	Rephrase-Suggest	100	100	100
	New Information	75	100	75
Created Procedure Student	Procedure Touch	**	**	**
	+	**	**	**
	-	**	**	**
	Feedback	**	**	**

Table 7 (Continued)

Category	Sub-Category	Judge AB	Judge AC	Judge BC
Afforded Procedure Teacher	Procedure Touch	**	**	**
	Feedback	100	100	100
Behavior	+	**	**	**
	-	**	**	**
	Warning	**	**	**
	Behavior Touch	**	**	**

** = no recordings were made; considered 100 percent agreement.
00 = one judge coded an interaction while another judge did not code an interaction; considered zero agreement.

this indicated 100 percent agreement. Zero agreement, indicated by the symbol 00, was interpreted as one interaction coded by one judge while another judge did not code an interaction (Lunt, 1972).

Collection of Data

Selection of School

The criteria used for the selection of a school to conduct this study were: the school should offer a number of different activities each period; each class should have at least 25-35 students; there should be at least three teachers of physical education. The make-up of the school's sex and racial balance was also considered. The school that met these criteria and that was recommended to the investigator was Allen Junior High School located in Greensboro, North Carolina.

After consultation with the principal of the school and the head of the Physical Education Department, the investigator sent a letter of introduction and explanation of the proposed study to the Assistant Superintendent of Public Instruction of the Greensboro Public School System. A proposal of the study was sent to the Director of Psychological Services as a result of a request for more details of the study. A short time later, the investigator received approval to conduct the study. A condition of approval was the requirement of furnishing a copy of the results of the study to the Division of Pupil Personnel Services.

Selection of Teachers, Classes, and Subjects

After a preliminary discussion with the Head of the Physical Education Department, four of the five teachers (three male and one female) were selected for the study. The investigator had a scheduled meeting with each teacher to explain the purpose of the study. Each teacher was told that the observers would be looking at the students' behavior in terms of the students' level of achievement. The teachers were not told that their behavior would also be recorded. After the preliminary explanation, the investigator collected the class rolls of all classes taught by the four teachers during the first four class periods of the day. Two teachers taught classes all four of the morning periods, one teacher taught three morning classes, and one teacher taught two morning classes.

Only the morning classes could be selected for the study because of the judges' previous commitments to graduate classes and

other engagements. From a total of 14 co-educational classes offered in the morning, four classes were selected for study. The activities being taught in these classes were volleyball, wrestling, speedball, and basketball. It was the investigator's intention to have four different activities, as well as four different teachers. Three of the classes (volleyball, wrestling, and basketball) were taught in the gymnasium and one activity, speedball, was taught out-of-doors. The numbers in classes ranged from 25 to 35. See Table 8.

After the classes were selected by the investigator, the teachers were then asked to rank their students (total group) in order of their physical achievement or skill potential. These rankings were made two days after the classes were selected. The rankings were used as the criterion measure of the teachers' expectations for their students' performance in physical education.

Ninety-six seventh-, eighth-, and ninth-grade students were involved in the study. Twenty-four students from each class were selected. The first 12 high achievers (starting with rank number one) and the first 12 low achievers (starting with the lowest number from the bottom) were selected for observation. Substitutes for each level (high and low) were also designated for observation on days when other subjects were not present. See Tables 9, 10, 11, and 12. The subjects, as well as the teachers, were given the impression that all of the students in each class were being observed and were part of the study. It was the original intent to use sex and race as criterion measures. The

Table 8

Schedule of Physical Education Activity Classes

Teacher	Activity	Class Number
	Period I - 8:30-9:15 a.m.	
*1	Volleyball	26
2	Wrestling	36
3	Physical fitness	21
4	Basketball	40
	Period II - 9:19-10:04 a.m.	
1	Physical fitness	36
*2	Wrestling	31
3	No assignment to physical education class	
4	Basketball	29
	Period III - 10:08-10:53 a.m.	
1	Physical fitness	23
2	Wrestling	20
*3	Speedball	25
4	Basketball	32
	Period IV - 10:57-11:42 a.m.	
1	No assignment to physical education class	
2	No assignment to physical education class	
3	No assignment to physical education class	
*4	Basketball	35

*Classes selected for observations.

Table 9

High and Low Achievers According to Rank
Seventh Grade Volleyball (26)

Judge A				Judge B				Judge C			
Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color
WM	H1	1	Y	WM	H2	1	R	WM	H3	1	B
BM	H4	2	Y	WM	H5	2	R	BM	H6	2	B
WM	H7	3	Y	BF	H8	3	R	WF	H9	3	B
BF	H10	7	Y	WF	H11	4	R	WM	H12	4	B
WF	L16	9	Y	WF	L15	5	R	BF	L17	5	B
BM	L19	10	Y	WF	L18	6	R	BF	L20	6	B
WM	L23	17	Y	WF	L21	7	R	WF	L22	7	B
BF	L26	18	Y	WF	L24	8	R	BF	L25	9	B

Two Substitutes - WM-H13, BF-L14

Race-Sex - WM = White Male, BM = Black Male, WF = White Female, BF = Black Female.
Rank - H = High, L = Low.
Color - Y = Yellow, R = Red, B = Blue.

Table 10

High and Low Achievers According to Rank
Eighth Grade Wrestling (30)

Judge A				Judge B				Judge C			
Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color
WM	H3	1	Y	WM	H2	1	R	WM	H1	1	B
WF	H4	2	Y	WF	H5	2	R	BM	H6	2	B
BM	H7	3	Y	WF	H8	3	R	WM	H9	3	B
WM	H11	7	Y	WM	H10	4	R	WF	H12	4	B
BF	L21	9	Y	WF	L20	5	R	BF	L19	5	B
WF	L24	10	Y	WF	L23	6	R	WF	L22	6	B
WM	L27	17	Y	BF	L26	7	R	BF	L25	7	B
BM	L30	18	Y	WM	L29	8	R	WM	L28	9	B
				Six Substitutes - BM-H13, BM-H14, BM-H15, WF-L16, WF-L17, WM-L18							

Race-Sex - WM = White Male, BM = Black Male, WF = White Female, BF = Black Female.
Rank - H = High, L = Low.
Color - Y = Yellow, R = Red, B = Blue.

Table 11

High and Low Achievers According to Rank
Ninth Grade Speedball (25)

Judge A				Judge B				Judge C			
Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color
BM	H1	1	Y	WM	H2	1	R	WM	H3	1	B
WM	H4	2	Y	WM	H5	2	R	WM	H6	2	B
WM	H7	3	Y	WM	H8	3	R	BM	H9	3	B
BM	H10	7	Y	BM	H11	4	R	WM	H12	4	B
WF	L15	9	Y	WF	L14	5	R	WM	L13	5	B
WF	L18	10	Y	WM	L17	6	R	WM	L16	6	B
BF	L21	17	Y	BF	L20	7	R	BF	L19	7	B
BF	L24	18	Y	BF	L23	8	R	BF	L22	9	B

One Substitute - L16

Race-Sex - WM = White Male, BM = Black Male, WF = White Female, BF = Black Female.
Rank - H = High, L = Low.
Color - Y = Yellow, R = Red, B = Blue.

Table 12

High and Low Achievers According to Rank
Ninth Grade Basketball (35)

Judge A				Judge B				Judge C			
Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color	Race Sex	Rank	Code Number	Code Color
WM	H1	1	Y	WM	H2	1	R	BM	H3	1	B
WF	H4	2	Y	WM	H5	2	R	WF	H6	2	B
WM	H7	3	Y	BM	H8	3	R	WM	H9	3	B
WF	H10	7	Y	WM	H11	4	R	BF	L12	4	B
WF	L26	9	Y	WF	L25	5	R	WF	L24	5	B
BF	L29	10	Y	BM	L28	6	R	BF	L27	6	B
OF	L31	17	Y	BF	L32	7	R	BF	L30	7	B
WF	L35	18	Y	BF	L34	8	R	WF	L33	9	B

11 Substitutes - WF-H13, WM-H14, BM-H15, WM-H16, WM-H17, WF-H18, WM-L19, WM-L20, WM-L21, WF-L22, WF-L23

Race-Sex - WM = White Male, BM = Black Male, WF = White Female, BF = Black Female.
Rank - H = High, L = Low.
Color - Y = Yellow, R = Red, B = Blue.

imbalance of classes in regard to these factors, however, made it impractical to use them for seeking clear interpretations of expectancy group differences.

Observation measures and dates. Three observers coded four different activity classes on six separate days within a two-week period. On the first day before the first class meeting, each judge was given four 3 x 5 cards for each of the four classes with the following information: (a) name of the activity, (b) grade, (c) names of the eight subjects, and (d) the code color and code number. On the first testing day, the investigator called each subject's name and gave out each pinnie so that each judge could place the subjects' names with their color and number. On subsequent days, the judges were responsible for giving out the pinnies to their subjects. All students in the class were given pinnies. In this way, neither teachers nor students knew who was being observed. Teachers and students were given the impression that the whole class was being observed. The three color codes used for observation were yellow, red, and blue. The two other colors used for the additional subjects not being observed were green and black and white striped.

One of the most important findings discovered in the training session was that due to the nature of certain physical education activities and the large amount of space used, the observers could not sit in one place to observe and code behavior. The observers were instructed to move with the teacher so that all interactions

could be seen and heard. This procedure was necessary to insure the validity of the category system, since no observation could be coded unless the observer heard and observed the interactions clearly.

In one activity class, wrestling, the judges sat in one place. The wrestling class was held in a small wrestling room and all of the interactions could be seen and heard easily by the judges. In the other three classes, it was necessary for the judges to follow the teacher.

After each class, the judges checked the coding sheets for any mistakes, made corrections, and discussed points of confusion with the investigator. After each class, the coding sheets were collected by the investigator.

Each activity class was 45 minutes in length. The actual observing time was approximately 30 minutes. Fifteen minutes were used for roll taking and giving out the pinnies to every subject. The total observer time for one class observed six times approximated 3 hours (6 classes x 30 minutes = 180 minutes). Testing dates are shown in Table 13.

Data Preparation

Brophy and Good (1969) indicated that a vast amount of raw data could accumulate if the entire observation system were used for any length of time. For this study, the investigator collected 136 coding sheets from the judges for six days of observation on 96 subjects. The raw data for each subject in each class were

Table 13

Observation Dates

Date	Time	Class
November 8	8:30 - 9:15	Volleyball
Monday	9:19 - 10:04	Wrestling
	10:08 - 10:53	Speedball
	10:57 - 11:42	Basketball
November 10	8:30 - 9:15	Volleyball
Wednesday	9:19 - 10:04	Wrestling
	10:08 - 10:53	Speedball
	10:57 - 11:42	Basketball
November 11	8:30 - 9:15	Volleyball
Thursday	10:57 - 11:42	Basketball
November 15	8:30 - 9:15	Volleyball
Monday	9:19 - 10:04	Wrestling
	10:08 - 10:53	Speedball
	10:53 - 11:42	Basketball
November 17	8:30 - 9:15	Volleyball
Wednesday	9:19 - 10:14	Wrestling
	10:08 - 10:53	Speedball
	10:57 - 11:42	Basketball
November 18	8:30 - 9:15	Volleyball
Thursday	9:19 - 10:04	Wrestling
	10:08 - 10:53	Speedball
	10:57 - 11:42	Basketball
November 19	9:19 - 10:04	Wrestling
Friday	10:08 - 10:53	Speedball

tabulated separately to form the basic measures derived from frequency (quantitative) and percentage (qualitative) scores. The raw data were transferred to five separate summary sheets for each subject. The number of summary sheets necessary for the analyses of data totalled 480 (five summary sheets for each of the 96 subjects). Twenty-four frequency measures and 32 percentage measures were derived from the coding.

Preparation of Summary Sheets

As a first step to preparation, two summary sheets were processed to aid in the tabulation of frequency scores. For the first sheet, "Teacher Feedback in Response Opportunities," four blank copies were required for each subject: one for feedback following correct answers, one for feedback following incorrect answers, one for feedback following no-response answers, and one for feedback following don't-know answers. The quality of the subject's answer was indicated by checking one of the four boxes at the top of the page. The coding sheet that the investigator used for observation was not designed for coding separate answers. The columns for coding the teacher's feedback were used for all quality answers (correct, incorrect, no response, and don't know). Therefore, four sheets indicating specifically the quality of answer of the subject's response were necessary, since knowledge of the quality of the subject's response was required before the codes could be interpreted.

A second summary sheet was used for recording the number of checks in columns (categories). The level-of-question category

and the procedural and behavioral dyadic contacts section were included in this summary sheet. These categories were simply summed and did not need to be sub-divided as in the feedback response opportunities' sheet. Samples of coding sheets and summary sheets can be found in the Appendix.

Frequency Measures (Quantitative)

Simple frequency counts were totalled for each column for each separate observation of every subject. These frequency totals formed the basic measures for the interpretation of data. Frequency counts were used to indicate how often events happened in single categories or combinations of categories. For example, the number of times a subject answers a question correctly would be interpreted as a single frequency score. The combination of frequency scores or categories would be interpreted as the number of times a subject answered (total number of correct, incorrect, no response, and don't know answers). Below are some examples of frequency measures taken for this study.

Single frequency scores

1. Total number of general touch
2. Total number of incidental touch
3. Total number of assist touch
4. Total number of procedural touch

Combination of frequency scores

The sum of questions 1-4 above (total number of all forms of touch).

Percentage Measures (Qualitative)

Frequency totals converted to percentage scores compares the quality of teacher-child interaction and the communication of expectations by the teacher to different individuals. For example, a subject may be touched by a teacher just as much as another subject in the same amount of time. The difference, however, may be in the quality of interaction. The percentage of general touch compared to the total number of touches may vary greatly from one subject to another. Some examples of the percentage measures taken for this study were as follows:

Praise (climate) and criticism of academic performance

1. Praise (climate) following correct answers over total answers.

2. Praise (climate) following wrong answers (includes don't know and no response) over total answers.

3. Criticism following right answers over total answers.

4. General touch following correct answers over total answers.

Analysis of Data

Analysis of variance was the statistic used to assess the effects of teacher expectations and class and their respective interactions on the obtained rankings, and to determine the effect of five different variables (climate, feedback, input, output, and touch) on high and low achievers. A computer analysis was used to obtain the scores at the alpha level of .05.

CHAPTER IV

ANALYSIS AND INTERPRETATION OF DATA

The purpose of this study was to identify specific and differential teacher behaviors that affect student behavior based on Rosenthal's Four Factor Theory (1974). Specific purposes included the following:

1. The identification of teachers' differential treatment of students according to the teachers' expectations (high or low).

2. The identification of students' different responses according to the teachers' expectations (high or low).

3. The addition of the factor, touch, as another possible identifying mechanism responsible for the Pygmalion Effect.

Four different physical education activity classes were selected for observational study. Teachers were asked to rank their students (total group in each class) in order of their physical achievement or skill potential. The rankings were used as the criterion measure of the teachers' expectations for their students' performance in physical education. Three judges, trained in the use of the Brophy and Good Interaction Analysis System (1969), observed 96 (24 students from each class) junior high school students on six separate days within a two-week period. Forty-eight of the students had been designated by their teachers as high achievers. The remaining 48 students

were designated as low achievers. The judges were unaware of the students' designations.

Summary sheets were developed for separate tabulations of the coded observations. Twenty-four frequency measures and 32 percentage measures were derived from the coding. Analyses of variance were performed on these measures to assess the effects of teacher expectations and class and their interactions on the obtained rankings, and to determine the effect of five different variables on high and low achievers.

The following five hypothetical statements which were presented in Chapter I were used as a guide for the analysis and interpretation of the data.

I. Teachers treat students differently according to the teachers' expectations (high or low expectations).

II. Teachers who expect superior achievement from their students treat them differently from students whom they consider to be low or inferior achievers in these four ways: (a) Climate (warmth or praise), (b) Feedback, (c) Output, and (d) Input (Rosenthal's Four Factory Theory, 1969).

III. Teachers touch their high-achieving students more than their low-achieving students.

IV. Teachers who touch their high-achieving students more also exhibit more warmth, feedback, output, and input.

V. Expectancy effects occur in physical education activity classes.

Each statement will be presented and discussed in the above order and will be analyzed and interpreted on the basis of quantitative (frequency) and qualitative (percentage) measures. The quantitative or frequency measures indicate how often interactions take place and show the objective or relative differences between groups. Data based on quantitative measures, however, cannot be construed solely as evidence that high or low expectancy groups are treated differently according to the teachers' expectations. Any significant evidence of quantitative measures indicates only that the possibility exists that teachers treat students differently according to their expectations.

To clearly and unequivocally assess teachers' expectancy effects, measures on absolute differences as well as relative differences must be taken. Qualitative or percentage measures give evidence of the quality of the teachers' behavior in absolute equal situations. Direct comparisons of high and low achievers in absolute equal circumstances must be measured to positively determine teacher expectancy effects that are due to differential teacher behaviors.

Statement I (one) will be analyzed and interpreted on the basis of quantitative measures taken on teachers' differential expectations of the high and low achieving groups. Statement II will be analyzed and separated into two sections. Three of the four variables (Climate, Feedback, and Output) in Statement II will be interpreted on the basis of quantitative (objective or relative differences) and qualitative (absolute differences)

measures. The fourth variable in Statement II (Input) will be interpreted on the basis of a qualitative measure. The variable, Touch, in Statements III and IV will be analyzed as one statement both quantitatively and qualitatively. Statement V will conclude with the summary of the analyses of data. Tables of Means and Analyses of Variance will accompany each statement. Because of the large number of tables accompanying each statement, only those statements that show significant effects will be presented for detailed analysis. All other tables will be included in the Appendix. The above procedure will be used to avoid confounding any of the significant effects found and to eliminate confusion regarding the interpretation of the data.

For the purpose of a clear and accurate analysis, the designated high and low expectancy groups (students ranked as high and low achievers by their teachers) will be referred to throughout the analyses as the high achievers and the low achievers.

It was not the purpose of this study to analyze differences between classes, but only to refer to the degree of variance across classes. Significant class interactions will be examined to determine the consistency of teacher expectancy effects.

Hypothetical Statement I

Teachers treat students differently according to the teachers' expectations (high and low expectations).

Data in Tables 14 and 15 show the objective or relative differences between the high and low expectancy groups.

Table 14

Analyses of Variance for Quantity and Type of Contacts

Variable	Source	Df	SS	Ms	F
Student-initiated questions	Level	1	.14	.14	.12
	Class	3	28.36	9.45	7.52**
	Level x Class	3	5.07	1.69	1.35
	Error	88	110.66	1.26	
	Total	95	144.24		
Response opportunities	Level	1	141.74	141.74	7.67**
	Class	3	450.41	150.13	8.13**
	Level x Class	3	44.74	14.91	0.81
	Error	88	1625.72	18.47	
	Total	95	2262.62		
Teacher-afforded procedural contacts	Level	1	34.98	34.98	6.16*
	Class	3	48.22	16.07	2.83*
	Level x Class	3	52.95	17.65	3.11*
	Error	88	499.74	5.67	
	Total	95	635.90		
Student-created procedural contacts	Level	1	.06	.06	.25
	Class	3	3.66	1.22	4.48**
	Level x Class	3	2.68	.89	3.28*
	Error	88	23.98	.27	
	Total	95	30.40		
Total procedural contacts	Level	1	38.14	38.14	6.79*
	Class	3	65.75	21.91	3.90*
	Level x Class	3	59.87	19.95	3.55*
	Error	88	494.23	5.61	
	Total	95	658.00		
Total behavioral contacts	Level	1	2.35	2.35	3.09
	Class	3	4.66	1.55	2.04
	Level x Class	3	7.82	2.60	3.42*
	Error	88	67.11	.76	
	Total	95	81.95		
Total dyadic contacts	Level	1	384.83	384.83	10.61**
	Class	3	741.23	247.07	6.91**
	Level x Class	3	143.74	47.91	1.32
	Error	88			
	Total	95			

* .05 level
** .01 level

Table 15

Mean Frequencies of Quantity and Type of Contacts Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Student-initiated questions	48	.68	48	.58	24	1.33	24	1.00	24	.08	24	.12
Response opportunities	48	9.87	48	7.50	24	11.04	24	10.70	24	6.62	24	6.37
Teacher-afforded procedural contacts	48	3.06	48	2.00	24	3.16	24	2.29	24	1.45	24	3.20
Student-created procedural contacts	48	.25	48	.18	24	.54	24	.20	24	.04	24	.08
Total procedural contacts	48	3.31	48	2.18	24	3.70	24	2.50	24	1.50	24	3.29
Total behavioral contacts	48	.64	48	.31	24	.83	24	.45	24	.41	24	.20
Total dyadic contacts	48	13.83	48	10.00	24	15.58	24	13.66	24	8.54	24	9.87

Different types of contacts and total number of contacts are observed as an indication that teachers treat students differently according to the teachers' expectations. Frequency measures were computed on seven substatements. These were: (1) student-initiated questions, (2) total response opportunities (comprised of direct, open, and call out questions), (3) teacher-afforded procedural contacts, (4) student-created procedural contacts, (5) total procedural contacts, (6) total behavioral contacts, and (7) total dyadic contacts (combination of total response opportunities, total procedural contacts, and total behavioral contacts).

Quantitative analysis. For four of the substatements, the difference between means for the high and low groups was statistically significant. This difference was significant at the .05 level of confidence for three of the substatements and significant at the .01 level of confidence for one substatement (see Tables 14 and 15, pages 92 and 93). Two of the substatements showed significant class interaction effects at the .05 level of confidence (see Figures 2 and 3).

1. Total response opportunities. The total number of direct, open, and call out questions which comprised the total response opportunity substatement showed a significant difference at the .05 level of confidence. High achievers were given more attention and were afforded more opportunities (i.e., were asked more questions) on an average of 9.8 times compared to the low achievers who were given opportunities to respond on an average of 7.5 times.

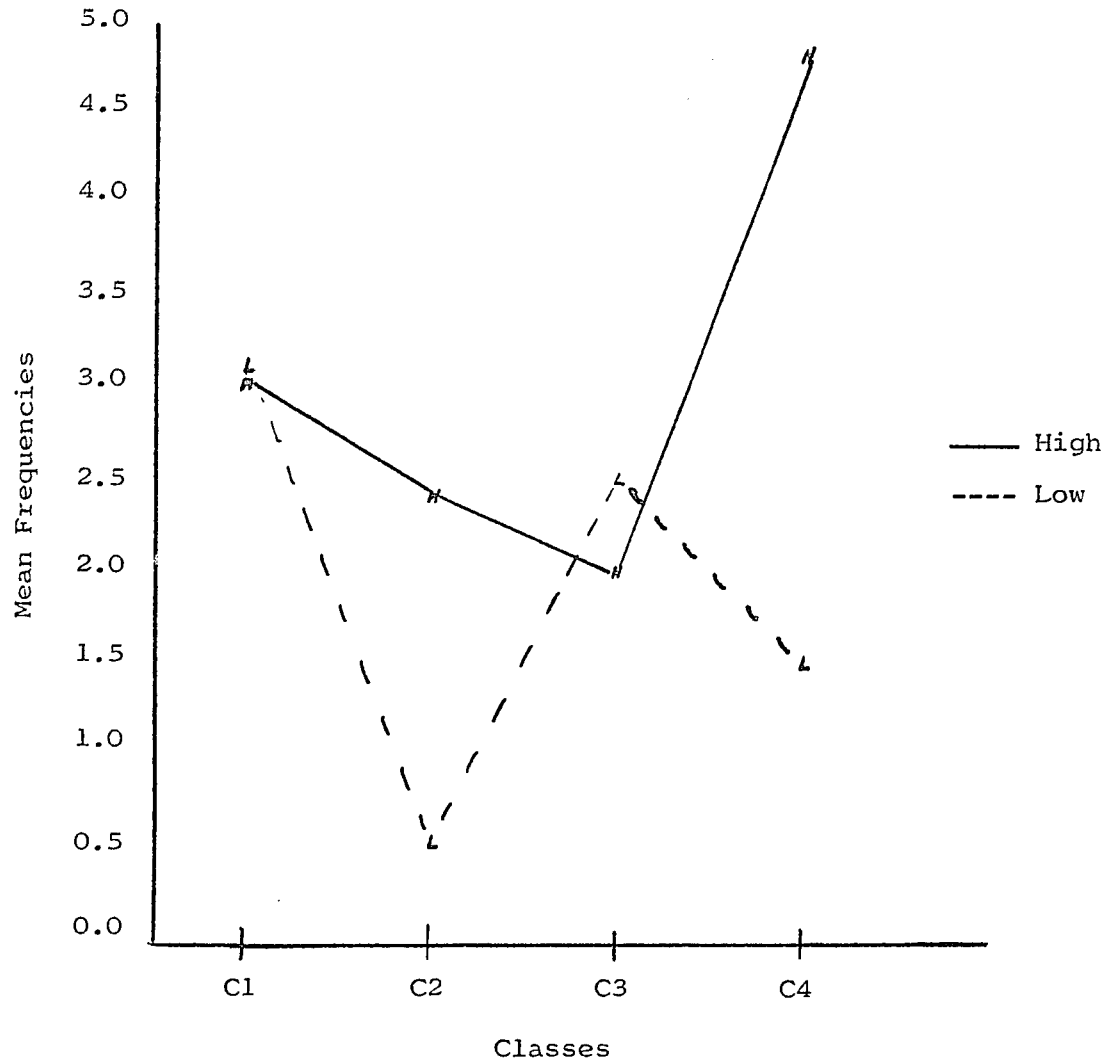


Figure 2

Significant Class Interaction for Teacher Afforded Procedural Contacts

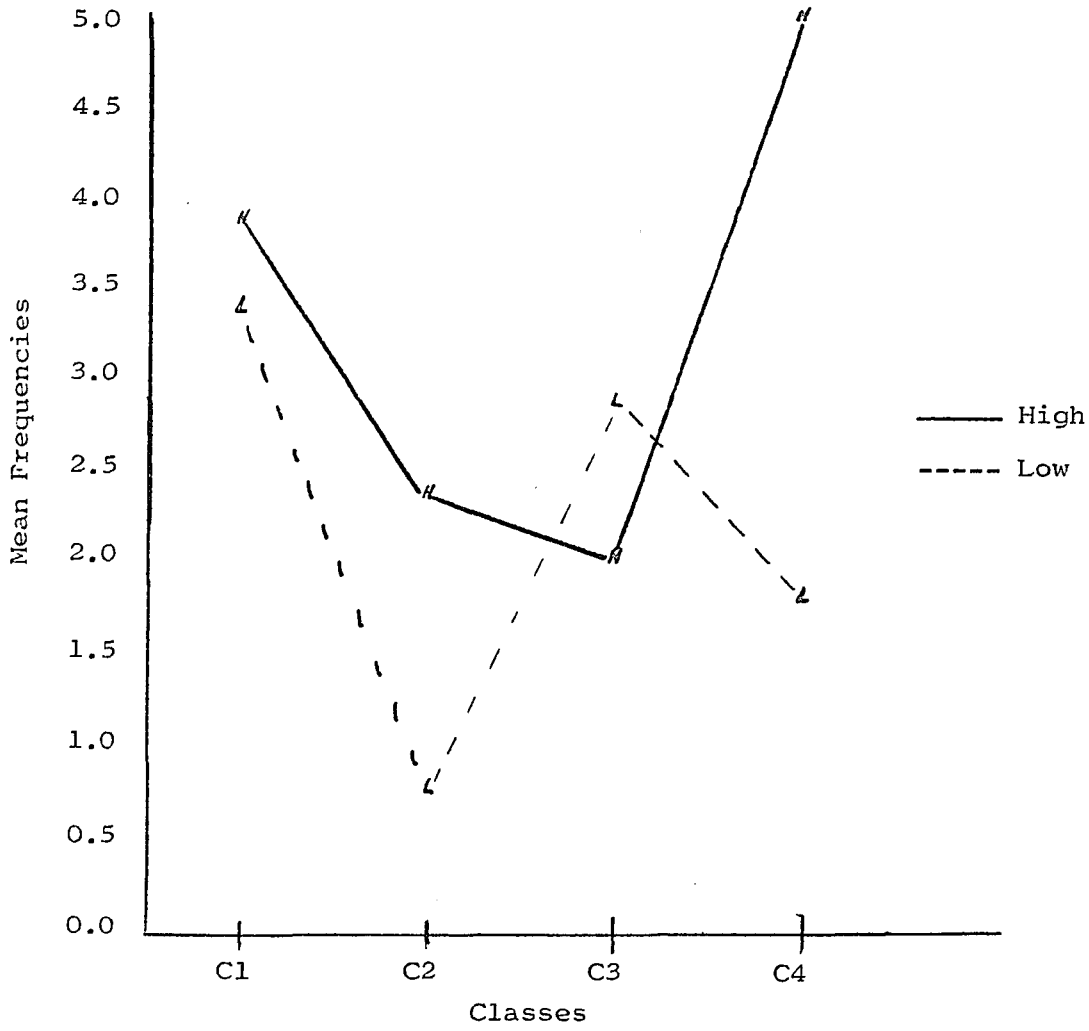


Figure 3

Significant Class Interaction for Total Procedural Contacts

2. Teacher afforded procedural contacts. The total number of times that teachers asked their high-achieving students to move equipment, or pass out equipment differed significantly at the .05 level of confidence. Teachers initiated more procedural contacts with high-achieving students than with the low-achieving students. A significant class interaction was also evident at the .05 level of confidence. The class x level (expectancy) interaction shown in Figure 2, page 95, indicates the degree of difference between classes. The direction of the effect was not consistent. Class two and Class four showed relatively larger differences between the high and the low achievers (i.e., more frequent interactions with the high achievers). The average of interactions for the high achievers in Class two was 2.41, and the average of interactions for the low achievers was .58. In Class four, the high achievers averaged 4.83 interactions and the low achievers averaged 1.58 interactions.

3. Total procedural contacts. Total procedural contacts comprised both student-created and teacher-afforded contacts. There was a significant difference at the .05 level of confidence. High-achieving students were contacted more times by the teacher to help with equipment and to perform other procedural activities than were the low-achieving students. There was no significant difference between the high and low achievers regarding student-created procedural contacts. Although the frequency of interactions

on this measure was minimal, the high-achieving students sought out the teacher to help with procedural activities (i.e., asking to help with equipment and arranging equipment) more than did the low-achieving students. The data for the combination of student-created and teacher-afforded procedural contacts produced a significant difference at the .05 level of confidence. A significant class interaction was also evident at the .05 level of confidence. The class x level (expectancy) interaction in Figure 3, page 96, showed that there were more frequent interactions with the high achievers than with the low achievers in two of the four classes. The mean for the high achievers in Classes two and four averaged 2.41 and 5.0 respectively, whereas the mean for the low achievers in Classes two and four averaged .58 and 1.58 interactions respectively.

4. Total dyadic contacts. There was a significant difference at the .01 level of confidence for the combination of total response opportunities, total procedural contacts, and total behavioral contacts. The total number of dyadic contacts that teachers made with the high achieving students averaged 13.8 times, whereas the total number of dyadic contacts made with the low achieving students averaged 10 times. There was no significant difference between the high and low achievers on total behavioral contacts because there were very few interactions concerning the students' conduct.

Discussion. The data for Statement I (Tables 14 and 15, pages 92 and 93) provide evidence that teachers do treat students differently according to the teachers' high or low expectations.

More attention was given to the high-achieving students in terms of frequency of contacts. High-achieving students were asked more questions and given more opportunities to respond 75% more of the time than were the low-achieving students. Teachers approached the high-achieving students 61% more of the time for procedural activities. As a result of the combination of the numbers of contacts made by the teachers in response opportunity interactions and procedural interactions (total dyadic contacts), high-achieving students were given the opportunity to interact and react with their teachers 72% more of the time than the low-achieving students.

The evidence in Figures 2 and 3 (pages 95 and 96) provide additional support that the teachers treat students differently according to the teachers' high or low expectations. The degree of the effect in two of the four classes (Class one and Class four) on teacher-afforded and total procedural contacts indicated a greater frequency of interactions with the high achievers than with the low achievers.

That teachers do afford more opportunities to their high-achieving students to perform does not unequivocally mean that these behaviors by the teacher are due to teacher expectancy effects. The data (quantitative measures) can be interpreted only as evidence indicating that there is a significant difference between the high and low achievers, and that these groups are treated differently. There is the possibility that the students' differential performances are the determinants in initiating the

teachers' expectations, which then result in teachers showing differential behaviors to particular groups of students.

Hypothetical Statement II: Climate

Teachers who expect superior achievement from their students treat them differently compared to their low or inferior achievers in four particular ways: Climate (warmth-praise), Feedback, Output, and Input.

The Climate Factor relates to the ways in which teachers create a warmer atmosphere for students. Friendliness, support and understanding, and gestures of approval are indicators that teachers are supportive of the Climate Factor.

The data for the Climate Factor in Statement II consists of both quantitative (frequency) and qualitative (percentage) measures.

The simple frequency measures indicate how often interactions take place and give evidence of the initial action of the teacher. For example, the first action by the teacher is to ask the student a question or to give a general task to the student. The qualitative (percentage) measures indicate the reaction of the teacher after the first contact is made with the student and takes into account the absolute differences between high-achieving students and low-achieving students. Percentage measures allow for direct comparison between groups in equal situations and give evidence of the quality of the teachers' behavior in those situations. For example, after the teacher asks the student a question (in either the high or low group), the teacher

waits for the answer and then makes an evaluative comment. These measures of the reactions of the teacher in equal situations provide the evidence of the teachers' expectations.

Quantitative measures were computed on one substatement. This substatement was the total number of times Climate (praise-warmth) was demonstrated by the teacher to a particular group. Qualitative measures were computed on five substatements. These were: (1) correct answers over total answers, (2) wrong answers over total answers, (3) climate following correct answers over total answers, (4) affirmation and climate of right answers over total right answers, and (5) climate following wrong answers (includes don't know and no response) over total answers.

Quantitative analysis. The one frequency measure supported the Climate Factor and showed a significant difference at the .01 level of confidence. (See Table 16.)

Table 16

Analyses of Variance of Quantitative Measures for Variable Climate

Variable	Source	Df	SS	Ms	F
Total number of times student praised or shown warmth by teacher	Level	1	282.71	282.71	11.32**
	Class	3	4.88	1.62	.07
	Level x Class	3	87.63	29.21	1.17
	Error	88	2198.72	24.98	
	Total	95	2573.94		

** .01 level

1. Total climate interactions. Table 17 provides evidence that teachers were consistent across classes and gave more praise, were more warm, and were generally more supportive to the high achievers than they were to the low achievers. The high-achieving students received more praise or were shown more warmth and support on the average of 9.3 times compared to the low achievers who experienced warmth and support from their teachers on the average of 5.8 times.

Table 17

Mean Frequencies of Quantitative Measures for Variable Climate Calculated According to Expectancy Group and Class

	High		Low		Class 1		Class 2		Class 3		Class 4	
Variable	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total number of times student praised or shown warmth by teacher	48	9.39	48	5.89	24	7.62	24	7.45	24	8.00	24	7.50

Qualitative analysis. For three of the five substatements, there were significant differences between groups at the .05 level of confidence. A fourth substatement did not show a significant difference. However, there was a significant class interaction effect at the .01 level of confidence for this substatement. (See Table 18 and Figure 4.)

Table 18

Analyses of Variance of Qualitative Measures for Variable Climate

Variable	Source	Df	SS	Ms	F
Correct answers over total answers	Level	1	2312.62	2312.62	4.49*
	Class	3	596.48	198.82	.39
	Level x Class	3	1068.35	356.11	.69
	Error		44774.20	514.64	
	Total		48751.65		
Wrong answers, no response, don't know, over total answers	Level	1	2500.83	2500.83	6.67*
	Class	3	223.41	74.47	0.20
	Level x Class	3	2513.19	837.73	2.23
	Error	81	30377.75	375.03	
	Total	88	35615.18		
Climate (praise) following correct answers over total answers	Level	1	755.28	755.28	2.75
	Class	3	5891.91	1963.97	7.15**
	Level x Class	3	3284.89	1094.96	3.99**
	Error	87	23880.79	274.49	
	Total	94	33812.89		
Climate (praise) and affirmation of correct answers over total answers	Level	1	16.46	16.46	4.89*
	Class	3	8.24	2.74	.82
	Level x Class	3	23.26	7.75	2.30
	Error	88	296.52	3.36	
	Total	95	344.48		
Climate (praise) following wrong answers over total answers (includes no response and don't know)	Level	1	136.47	136.47	1.22
	Class	3	497.63	165.87	1.48
	Level x Class	3	574.45	191.48	1.71
	Error	87	9718.40	111.70	
	Total	94	10926.97		

* .05 level
** .01 level

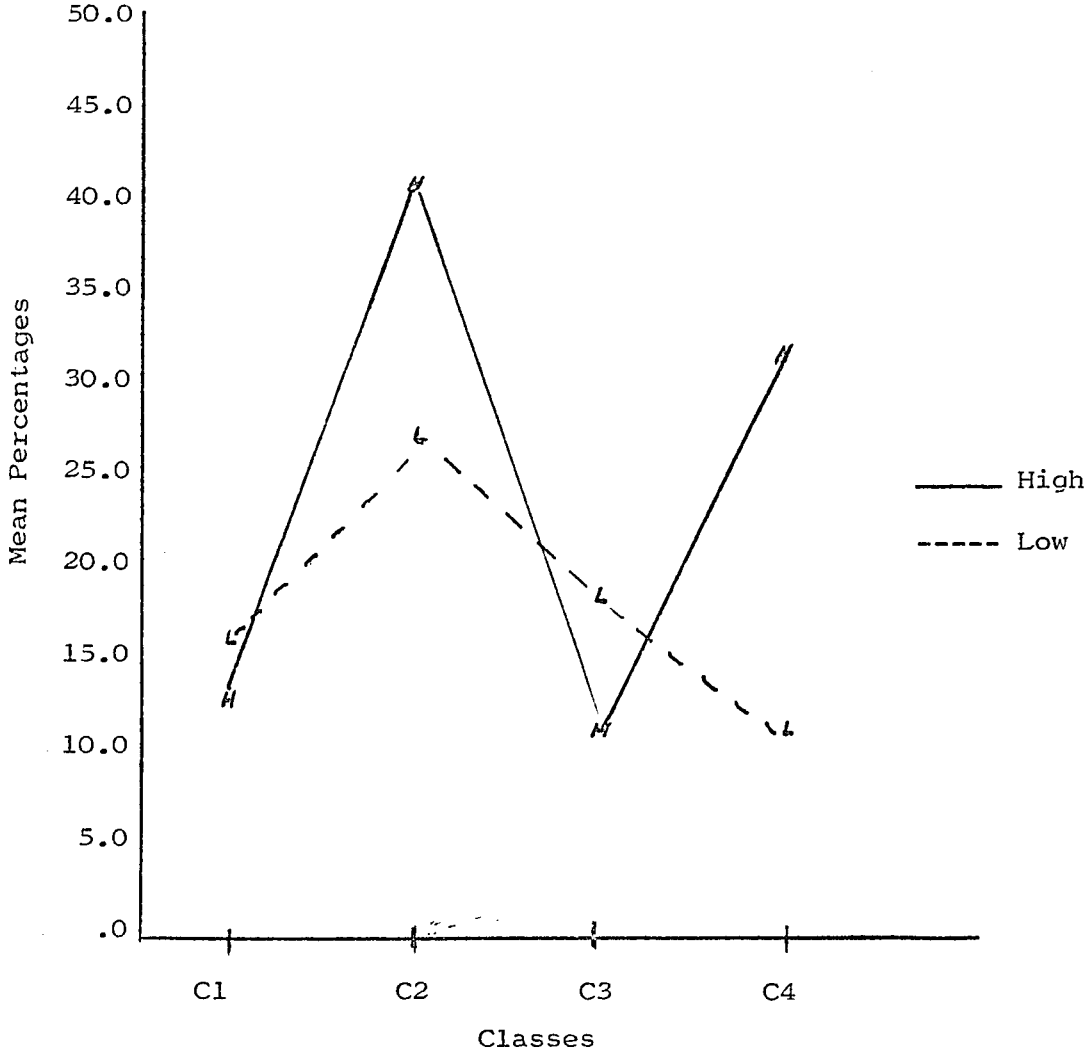


Figure 4

Significant Class Interaction for Climate Factor: Praise Following Correct Answers Over Total Answers

1. Correct answers over total answers. The data indicated that high-achieving students answered more questions directed to them by their teachers correctly than did the low-achieving students. A significant difference at the .05 level of confidence indicated that the high achievers gave more correct answers over their total answers than did the low achieving group. This measure was taken to see if there was a difference in the amount of praise for students who answered more questions of their total questions correctly, than for students who did not answer more questions correctly. From the total number of answers given, the high achievers answered questions correctly 59.8% of the time compared to the low achievers who answered questions correctly 49.6% of the time.

2. Wrong answers over total answers. A significant difference at the .05 level of confidence indicated that high-achieving students did not give as many wrong answers as did the low achievers. From the total number of answers given, the low achievers answered incorrectly 53.7% of the time, whereas the high achievers gave wrong answers only 43.1% of the time.

3. Climate following correct answers over total answers. There was no significant difference to show that teachers exhibited more warmth or praise to students who answered more questions correctly. Teachers gave warmth and support 24% of the time to the high achievers and gave support to the low achievers 18.3% of the time. The absence of a significant level effect indicates that teachers were not consistent across classes, because there

was a significant class interaction effect at the .05 level of confidence. Figure 4, page 104, shows the averages in Classes two and four. In Class two, the high achievers received more praise and warmth when they answered questions correctly 40.4% of the time. The low achievers in the same class received praise 27.1% of the time when they answered questions correctly. In Class four, the high achievers received praise from the teacher 31.8% of the time when they answered questions correctly, whereas the low achievers were given praise and support on the average of 11.1% of the time when they answered questions correctly.

4. Affirmation and climate of right answers over total right answers. There was a significant difference at the .05 level of confidence indicating that teachers gave more warmth and support to high achievers when they answered correctly. Low achievers did not receive as much praise with affirmation when they answered the teachers questions correctly or when they made the appropriate movement responses. Teachers affirmed the correct responses to the high achievers and at the same time were more supportive of them. High achievers were given affirmation coupled with praise 2.54% of the time. The low achievers received the same kind of treatment 1.70% of the time. (See Table 19.)

5. Climate following wrong answers (includes don't know and no response) over total answers. There was no significant difference between the high and low groups when comparisons were made as to the amount of praise following wrong answers over the total number of answers. The frequency of interactions was small.

Table 19

Mean Percentages of Qualitative Measures for Variable Climate Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Correct answers over total answers	47	59.86	48	49.66	24	54.31	24	57.95	23	50.82	24	55.58
Wrong answers, no responses, don't know, over total answers	45	43.10	44	53.70	23	47.66	21	48.04	22	51.41	23	46.35
Climate (praise) following correct answers over total answers	48	24.00	47	18.36	24	14.53	24	33.80	23	14.71	24	21.51
Climate (praise) and affirmation of correct answers over total answers	48	2.54	48	1.70	24	2.04	24	2.50	24	1.70	24	2.25
Climate (praise) following wrong answers over total answers (including no response and don't know)	48	3.40	47	1.01	24	.00	24	5.97	23	1.17	24	1.69

Discussion. The data for the Climate Factor in Statement II provide evidence indicating that teachers treat their high-achieving students more warmly than they treat their low-achieving students. The significant difference, however, cannot be considered as an unequivocal conclusion that teachers communicate their differential expectations by manifesting differential behaviors. The quantitative measure or the frequency of praise and warmth by the teachers can be interpreted only as a causal factor and attributed only to objective differences between groups. The possibility exists that the effect may be reversed whereby the students create the expectation for the teachers and influence teachers' differential behavior. The evidence clearly indicates that low-achieving students did not receive equal amounts of praise and warmth suggesting that this effect is due to differential teacher expectations. On the other hand, the possibility exists that this significant difference may not be due to differential teacher expectations but to differential student performance leading to the creation of teacher expectations.

When looking at absolute differences between the high achievers and the low achievers, the data strongly suggest that teachers do communicate their differential expectations through their own behavior. The evidence indicates that when teachers expected students to perform better, the teachers treated those students more warmly and gave more support (both verbally and non-verbally) than they treated those students of whom they expected less. For example, high-achieving students answered more questions correctly

and as a result received more affirmation coupled with praise for their correct performance than did the low achievers. The level of performance of the low achievers was less than the performance of the high achievers. Low achievers answered more questions incorrectly and although the difference was not significant, teachers did praise the high achievers more than they praised the low achievers when questions or responses were incorrect. From the above evidence, it is clear that teachers display differential behaviors to the high and low achievers under equal situations.

Hypothetical Statement II: Feedback

Teachers who expect superior achievement from their students treat them differently compared to their low or inferior achievers in four particular ways: Climate (warmth-praise), Feedback, Output, and Input.

Both Feedback and Climate consist of praise and support. The distinctive factor, however, is the degree to which warmth or praise is dependent upon the kind of response that the student has made. For example, warmth or praise that is specifically related to a correct or desired response indicated by the teacher can be considered supportive of the Feedback Factor (Rosenthal, 1974).

The Feedback Factor also relates to appropriate feedback given by the teacher when a student makes an incorrect response. This type of active teaching can be in the form of supplying the answer, partially correcting the answer, asking another

student to help with the answer or giving clues to the student to reshape the response.

Data for the Feedback Factor in Statement II consists of both quantitative (frequency) and qualitative (percentage) measures. Frequency measures for the Feedback Factor were computed on three substatements. These were: (1) total correct answers, (2) total wrong answers, and (3) sustained feedback (sums of repeat, rephrase, and new questions). Qualitative measures were computed on six substatements. These measures were: (1) affirmation and climate of right answers over total right answers, (2) negations (including criticisms) following wrong answers over total wrong answers, (3) criticism following right answers over total answers, (4) number of no feedback over total correct responses, (5) give answer over total incorrect responses, and (6) failure followed with sustained feedback over total failures.

Quantitative analysis. One of the three substatements showed a significant difference at the .01 level of confidence. (See Table 20.)

1. Total correct answers. A significant difference at the .01 level of confidence indicated that high-achieving students had a total of more correct answers than did the low-achieving students. High-achieving students answered more questions correctly on the average of 6.91 times while the low achievers answered questions correctly on the average of 4.91 times. (See Table 21.)

Table 20

Analyses of Variance of Quantitative Measures for Variable Feedback

Variable	Source	Df	SS	Ms	F
Amount of sustained feedback	Level	1	.68	.68	.06
	Class	3	165.24	55.08	4.69**
	Level x Class	3	31.19	10.39	.88
	Error	88	1034.61	11.75	
	Total	95	1231.72		
Total number of correct answers	Level	1	100.42	100.42	7.08**
	Class	3	128.34	42.78	3.02*
	Level x Class	3	17.16	5.72	.40
	Error	88	1247.40	14.17	
	Total	95	1493.32		
Total number of wrong answers	Level	1	.30	.30	.02
	Class	3	107.28	35.76	2.90*
	Level x Class	3	63.70	21.23	1.72
	Error	81	998.46	12.32	
	Total	88	1169.74		

* .05 level
** .01 level

Table 21

Mean Frequencies of Quantitative Measures for Variable Feedback Calculated According to Expectancy Group and Class

	High		Low		Class 1		Class 2		Class 3		Class 4	
Variable	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total number of correct answers	48	6.91	48	4.91	24	7.79	24	4.54	24	5.91	24	5.41
Total number of wrong answers	44	5.81	45	5.80	23	7.04	21	4.42	22	6.68	23	5.00
Amount of sustained feedback	48	3.50	48	3.39	24	4.75	24	1.66	24	2.70	24	4.66

2. Total wrong answers. There was no significant difference for this measure. Both high achievers and low achievers answered incorrectly approximately the same number of times.

3. Sustained feedback (includes repeat, rephrase and new questions). There was no significant difference to show that teachers repeated or rephrased their questions more to one group of students over the other. High achievers received feedback on the average of 3.50 times and the low achievers received sustained feedback on the average of 3.39 times (See Table 21, page 111).

Qualitative analysis. The data for one of the six substatements indicated a significant difference at the .05 level of confidence. Data from the other five substatements did not provide any significant differences. (See Table 22.)

1. Affirmation and climate of right answers over total right answers. Feedback specifically related to a desired response by the teacher was significantly different at the .05 level of confidence. When high-achieving students answered a question correctly, teachers evaluated them affirmatively and at the same time showed warmth and support specifically in relation to the particular answer. On the other hand, when low achievers answered correctly, the teachers did not give additional warmth or support to them even though the response was the correct one. The high achievers received feedback on the average of 2.54% of the time and the low achievers experienced feedback from their teacher 1.70% of the time. (See Table 23.)

Table 22

Analyses of Variance of Qualitative Measures for Variable Feedback

Variable	Source	Df	SS	Ms	F
Climate (praise) and affirmation of right answers of total right answers	Level	1	16.46	16.46	4.89*
	Class	3	8.24	2.74	.82
	Level x Class	3	23.26	7.75	2.30
	Error	88	296.52	3.36	
	Total	95	344.48		
Negations (including criticism) following wrong answers over total wrong answers	Level	1	470.15	470.15	.30
	Class	3	38068.31	12689.43	8.19**
	Level x Class	3	4032.22	1344.07	.87
	Error	81	125510.04	1549.50	
	Total	88	168080.72		
Criticism following right answers over total answers	Level	1	3.02	3.02	1.91
	Class	3	3.61	1.20	.76
	Level x Class	3	3.60	1.20	.76
	Error	87	137.67	1.58	
	Total	94	147.90		
Number of No feedback over total correct responses	Level	1	432.67	432.67	1.45
	Class	3	11670.25	3890.08	13.00**
	Level x Class	3	899.86	299.95	1.00
	Error	84	25128.07	299.14	
	Total	91	38130.85		
Give answers over total wrong answers	Level	1	511.66	511.66	.45
	Class	3	3186.99	1062.33	.93
	Level x Class	3	667.43	222.47	.20
	Error	81	92113.48	1137.20	
	Total	88	96479.56		
Failure followed with sustained feedback over total failures	Level	1	16.09	16.09	.03
	Class	3	12948.29	4316.09	7.97**
	Level x Class	3	1806.23	602.07	1.11
	Error	59	31967.01	541.81	
	Total	66	46737.62		

* .05 level
** .01 level

Table 23

Mean Percentages of Qualitative Measures for Variable Feedback Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Climate (praise) and affirmation of right answers over total right answers	48	2.54	48	1.70	24	2.04	24	2.50	24	1.70	24	2.25
Negations (including criticisms) following wrong answers over total wrong answers	45	36.48	44	41.08	23	6.67	21	46.21	22	63.46	23	40.41
Criticism following right answers over total answers	48	.35	47	.00	24	.00	24	.00	23	.43	24	.29
Number of no feedback over total right answers	47	14.40	45	18.74	24	33.74	24	6.62	22	7.01	22	18.05
Give answers over total wrong answers	45	52.52	44	47.73	23	51.89	21	59.53	22	46.24	23	43.59
Failure followed with sustained feedback over total failures	31	49.39	36	50.32	22	53.66	11	43.64	17	30.61	17	68.46

2. Negations (including criticisms) following wrong answers over total wrong answers. There was no significant difference to indicate that teachers gave more or less criticism to either group when questions were answered incorrectly.

3. Criticism following correct answers over total answers. There was no evidence to show that teachers gave more or less criticism to either expectancy group. The frequency of this interaction was negligible.

4. Number of no feedback over total correct responses. There was no significant difference to show that teachers were not actively giving feedback to the high or the low achievers. Low achievers were not given any feedback 18.7% of the time, whereas the high achievers did not receive feedback on the correct responses 14.4% of the time.

5. Give answer over total incorrect responses. There was no significant difference to show that teachers gave appropriate feedback for an incorrect response to either the high or the low achievers. High achievers were given the answer to an incorrect response 52.2% of the time. The low achievers were given an answer for an incorrect response 47.7% of the time.

6. Failure followed with sustained feedback over total failures. There was no significant difference to indicate that teachers helped high-achieving students more than low-achieving students when the student had failed.

Discussion. The data in Statement II does not provide strong support for the Feedback Factor. The frequency measure on sustained feedback did not show a significant difference. Teachers

gave just as much help to both groups even though the level of performance of the high achievers was significantly better.

The analysis of the absolute differences between groups indicated that there were significant differences in the amounts of affirmation coupled with praise. These data suggest that teachers do communicate their differential expectations through differential behaviors which, in turn, leads to differential performance expectations. However, the five other substatements did not show any significant differences and as a result there was a minimal amount of evidence to indicate that teachers give more differentiated feedback (terminal or sustained) to high-achieving students. Only weak support can be given to the Feedback Factor to confirm teacher expectancies.

Hypothetical Statement II: Output

Teachers who expect superior achievement from their students treat them differently compared to their low or inferior achievers in four particular ways: Climate (warmth-praise), Feedback, Output, and Input.

The Output Factor operates closely with the Feedback Factor. Teachers encourage greater responsiveness by calling more often on students. Students who are given more opportunities to respond are more likely to be given more feedback when the response is appropriate or accurate. Greater demands are imposed on students, and they are given more opportunities to demonstrate their abilities and skills. As in the case of the Feedback Factor, teachers are

more apt to shape an answer or extend a response by giving suggestions or making corrections.

Data for the Output Factor in Statement II consists of both quantitative (frequency) and qualitative (percentage) measures. Frequency measures for the Output Factor were taken on six substatements. These were: (1) total number of direct questions, (2) total number of open questions, (3) total number of call out questions, (4) total response opportunities, (5) total amount of sustained feedback, and (6) total number of answers. Percentage measures were taken on eleven substatements which were: (1) direct questions over total response opportunities, (2) open questions over total response opportunities, (3) call out questions over total response opportunities, (5) process questions over total questions, (6) product questions over total questions, (7) correct answers over total answers, (8) wrong answers, no response, and don't know answers over total answers, (9) rephrase following right answers over total right answers, (10) repeat over repeat plus rephrase plus new question following failure, and (11) failure followed with sustained feedback over total failures.

Quantitative analysis. For three of the six substatements, the difference between means for the high and low groups was statistically significant at the .05 level of confidence. (See Table 24.)

1. Total number of direct questions. There was a significant difference at the .05 level of confidence in the number of direct questions posed to the high achievers as compared to the low

Table 24

Analyses of Variance of Quantitative Measures for Variable Output

Variable	Source	Df	SS	Ms	F
Direct questions	Level	1	90.69	90.69	5.72*
	Class	3	178.03	59.34	3.74*
	Level x Class	3	31.93	10.64	.67
	Error	88	1395.28	15.85	
	Total	95	1695.93		
Open questions	Level	1	.19	.19	.29
	Class	3	11.63	3.87	5.64**
	Level x Class	3	.61	.30	.30
	Error	88	60.54	.68	
	Total	95	72.97		
Call out questions	Level	1	1.61	1.61	5.31*
	Class	3	2.50	.83	2.74*
	Level x Class	3	.58	.19	.64
	Error	88	26.70	.30	
	Total	95	31.39		
Total response opportunities	Level	1	141.74	141.74	7.67**
	Class	3	450.41	150.13	8.13**
	Level x Class	3	44.74	14.91	.81
	Error	88	1625.72	18.47	
	Total	95	2262.61		
Sustained feedback	Level	1	.68	.68	.06
	Class	3	165.24	55.08	4.69**
	Level x Class	3	31.19	10.39	.88
	Error	88	1034.61	11.75	
	Total	95	1231.72		
Total answers	Level	1	100.89	100.89	2.79
	Class	3	480.79	160.25	4.42**
	Level x Class	3	76.96	25.65	.71
	Error		3187.57	36.22	
	Total		3846.21		

* .05 level
** .01 level

achievers. Teachers made an obvious attempt to call on students of whom they expected more. High achievers were directly approached on the average of 8.27 times while the low achievers were directly approached on the average of 6.43 times. (See Table 25.)

2. Total number of open questions. There was no significant difference for this measure. The frequency of interactions was very small.

3. Total number of call out questions. There was a significant difference at the .05 level of confidence for call out questions; however, the frequency of interactions was minimal. Teachers recognized the high achievers who called out answers more than they recognized the low achievers when they called out the answer to a question posed to the class.

4. Total response opportunities. A significant difference at the .05 level of confidence showed that teachers gave more opportunities for the high-achieving students to respond. The high achievers were given opportunities to respond on the average of 9.87 times. The low achievers made responses when given the opportunity on the average of 7.50 times.

5. Total amount of sustained feedback (includes repeat, rephrase and new questions). There was no significant difference showing that teachers repeated or rephrased their questions more to the high achievers than to the low achievers.

6. Total number of answers. Although there was not a significant difference for total number of answers, the effect was highly suggestive in the direction of the high achievers. This measure

Table 25

Mean Frequencies of Quantitative Measures of Variable Output Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Direct questions	48	8.27	48	6.43	24	9.00	24	6.29	24	8.41	24	5.70
Open questions	48	.37	48	0.27	24	.20	24	.12	24	.91	24	.04
Call out questions	48	.41	48	0.14	24	.50	24	.08	24	.37	24	.16
Total response opportunities	48	9.87	48	7.50	24	11.04	24	6.62	24	10.70	24	6.37
Sustained feedback	48	3.50	48	3.39	24	4.75	24	1.66	24	2.70	24	4.66
Total answers	48	12.25	48	10.35	24	14.54	24	8.41	24	12.04	24	10.20

corresponds with the number of direct questions and total response opportunities relating to amount of attention afforded a particular group.

Qualitative analysis. Two of the 11 substatements showed significant differences at the .05 level of confidence. However, the other nine substatements which would give added and stronger support for teacher expectations did not indicate any significant differences. (See Table 26.)

The two substatements which were statistically significant (correct answers over total answers and wrong answers over total answers) coincided with the amount of attention and opportunities given to the high achievers compared to the attention and opportunities given to the low achievers. (See Tables 26 and 27.)

Discussion. The data in Statement II does not provide strong support for teacher expectations relating to the Output Factor. Teachers did initiate more interactions by directly asking questions of the high-achieving students thus giving these students more opportunities to respond. The evidence suggests that teachers did pay more attention to the high achievers. Consequently, the high-achieving students' level of performance coincided with the teachers' expectations. The high achievers received more attention, were given more opportunities to respond, and responded in the appropriate way by answering more questions correctly. Although teachers were more attentive to the high achievers, there was no difference in the amount of reinforcement or sustained feedback between the high achievers and the low

Table 26

Analyses of Variance of Qualitative Measures for Variable Output

Variable	Source	Df	SS	Ms	F
Direct questions over response opportunities	Level	1	69.32	69.32	.16
	Class	3	3441.46	1147.15	2.62
	Level x Class	3	196.21	65.40	.15
	Error	87	38162.05	438.64	
	Total	94	41869.04		
Open questions over response opportunities	Level	1	2.89	2.89	.04
	Class	3	942.65	314.21	3.80*
	Level x Class	3	65.48	21.82	.26
	Error	87	7200.41	82.76	
	Total	94	8211.43		
Call out questions over response opportunities	Level	1	51.42	51.42	2.39
	Class	3	110.35	36.78	1.71
	Level x Class	3	66.79	22.26	1.03
	Error	87	1875.54	21.55	
	Total	94	2104.10		
Student initiated response opportunities over total response opportunities	Level	1	.28	.28	.00
	Class	3	1832.98	610.99	6.45**
	Level x Class	3	127.27	42.42	.45
	Error	87	8235.91	94.66	
	Total	94	10196.44		
Process questions over total questions	Level	1	.06	.06	.09
	Class	3	3.82	1.27	1.92
	Level x Class	3	.32	.10	.16
	Error	87	57.85	.66	
	Total	94	62.05		
Product questions over total questions	Level	1	16.75	16.75	.14
	Class	3	282.67	94.22	.77
	Level x Class	3	169.63	56.54	.46
	Error	87	10630.60	122.19	
	Total	94	11099.65		
Correct answers over total answers	Level	1	2312.62	2312.62	4.49*
	Class	3	596.48	198.82	.39
	Level x Class	3	1068.35	356.11	.69
	Error	87	44774.20	514.64	
	Total	94	48751.65		

Table 26 (continued)

Variable	Source	Df	SS	Ms	F
Wrong answers no response, don't know, over total answers	Level	1	2500.83	2500.83	6.67*
	Class	3	223.41	74.77	0.20
	Level x Class	3	2513.19	837.73	2.23
	Error	81	30377.75	375.03	
	Total	88	35615.18		
Rephrase following right answers over total right answers	Level	1	20.48	20.48	.11
	Class	3	3192.94	1064.31	5.54**
	Level x Class	3	284.90	94.96	.49
	Error	84	16145.29	192.20	
	Total	91	19643.61		
Repeat over repeat plus rephrase plus new question following failure	Level	1	.38	.38	
	Class	3	1273.54	424.51	.00
	Level x Class	3	3728.71	1242.90	.96
	Error	60	26638.98	443.98	2.80*
	Total	67	31641.61		
Failure followed with sustained feedback over total failures	Level	1	16.09	16.09	.03
	Class	3	12948.29	4316.09	7.97**
	Level x Class	3	1806.23	602.07	1.11
	Error	59	31967.01	541.81	
	Total	66	46737.62		

* .05 level
** .01 level

Table 27

Mean Percentages of Qualitative Measures of Variable Output Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Direct questions over response opportunities	48	85.31	47	87.02	24	84.35	23	94.66	24	77.91	24	88.06
Open questions over response opportunities	48	3.66	47	3.31	24	1.94	23	2.50	24	8.79	24	.69
Call out questions over response opportunities	48	3.11	47	1.64	24	4.00	23	1.26	24	2.78	24	1.44
Student initiated response opportunities over total response opportunities	48	5.81	47	5.70	24	9.69	23	1.20	24	10.50	24	1.46
Process questions over total questions	48	.09	47	.14	24	.45	23	.00	24	.00	24	.00
Product questions over total questions	48	58.21	47	59.05	24	57.85	23	56.12	24	60.18	24	60.26
Correct answers over total answers	48	59.59	47	49.72	24	54.31	24	57.95	23	50.82	24	55.58

Table 27 (continued)

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Wrong answers, no response, don't know over total answers	45	43.10	44	53.70	23	47.66	21	48.04	22	51.41	23	46.35
Rephrase following right answers over total right answers	47	9.08	45	8.13	24	7.95	24	4.70	22	3.52	22	18.71
Repeat over repeat plus rephrase plus new question following failure	32	9.06	36	9.21	23	4.05	11	15.45	17	12.74	17	8.33
Failure followed with sustained feedback over total failures	31	49.39	36	50.38	22	53.66	11	43.64	17	30.61	17	68.46

achievers. High achievers did respond more, and not only answered more questions, but also answered these questions correctly. However, the amount of attention afforded by the teacher to the high achievers did not result in increased reinforcement of quality performance. These percentage measures of correct answers over total answers, and wrong answers over total answers, do suggest that teachers encourage students to respond in appropriate ways which would confirm teachers' expectations. However, the other nine substatements did not indicate that teachers discriminated in favor of the highs in demanding and reinforcing quality performance. Because of the minimal amount of evidence collected, only limited support can be given to the Output Factor for confirming teacher expectations.

Hypothetical Statement II: Input

Teachers who expect superior achievement from their students treat them differently compared to their low or inferior achievers in four particular ways: Climate (warmth-praise), Feedback, Output, and Input.

The Input Factor describes the amount of new material taught or amount of new information given to students. Data for the Input Factor in Statement II consists of one qualitative measure. No frequency measures were taken on the amounts of new material taught to students.

Qualitative analysis. The one qualitative measure did not support the Input Factor.

1. New questions following right answers over total answers. There was no significant evidence to indicate that teachers taught more new material to students of whom more was expected. (See Tables 28 and 29 in Appendix B.)

Hypothetical Statements III and IV: Touch

Teachers touch their high achieving students more than their low achieving students and exhibit more Climate, Feedback, Output, and Input to the high achievers.

Quantitative analysis. There was no evidence to suggest that teachers touched their high achievers more than they touched their low achievers. There were few interactions, and the data indicated that teachers did not touch students very many times. The frequency of interactions was minimal across classes.

Qualitative analysis. The frequency of interactions was minimal since teachers did not touch students many times. There was not sufficient data to support the hypothesis that when teachers do touch their high-achieving students more than they touch low-achieving students, they also exhibit more Climate, Feedback, Output, and Input. (See Tables 30 through 39 in Appendix B for quantitative and qualitative measures on the variable Touch.)

Hypothetical Statement V

Expectancy effects occur in physical education activity classes.

Evidence from the data presented in Tables 14 through 27 shows that expectancy effects occur in physical education classes. This evidence is summarized as follows:

1. Quantity and types of contacts. High-achieving students were asked more questions and given more opportunities to respond 75% more of the time than the low-achieving students. Teachers approached the high achievers 61% more of the time for procedural activities. High-achieving students were given the opportunity to interact and react with their teachers 71% more of the time than the low-achieving students for total dyadic contacts. The data indicated that there was a significant difference between the high and the low expectancy groups and that these groups were treated differently according to the teachers expectations. The data, however, do not specifically indicate whether the expectations were determined by the students or by the teachers.

2. Climate. The data for the Climate variable provide evidence that teachers treat their high-achieving students more warmly than they treat their low-achieving students. Teachers gave praise to their high-achieving students 62% more than they gave praise to their low achievers. The frequency measure indicates that the amount of praise and warmth can be interpreted only as a causal factor and attributed only to objective differences.

The measures on direct comparisons of both groups under equal circumstances indicated that when teachers expected students to perform better, the teachers treated those students more warmly and gave more support (both verbally and non-verbally) than they did to those students of whom they expected less. High-achieving students answered more questions correctly and as a result received

more evaluative comments (affirmation and praise) for their correct performance than did the low achievers. Teachers also praised the high achievers more than they praised the low achievers when responses were incorrect. The evidence clearly indicates that teachers display differential behaviors to the high and low achievers under equal situations.

3. Feedback. The data for the variable Feedback indicated some support for teacher expectations, but the evidence is limited. Although the level of performance was significantly greater for the high achievers than for the low achievers, there was not enough evidence to suggest that teachers gave more sustained feedback to the high achievers. However, there were significant differences in the amount of affirmation coupled with praise which suggests that teachers do reward their high achievers for the desired or correct responses expected from the teacher. The low achievers were not praised for their good performance as much as the high achievers. Five of the substatements did not show significant differences indicating that more evidence is needed to support the Feedback Factor. Further work is needed to confirm teacher expectancies relating to the Feedback variable.

4. Output. Strong support was not provided for the Output Factor. The evidence does significantly suggest that high achievers did receive more attention and were given more opportunities to respond more than the low achievers. The evidence, however, did not show a significant difference in the amount of reinforcement given by the teachers. However, when the highs responded with

correct answers, the teachers rephrased the question or made suggestions providing a second response opportunity for the high achievers more than they did of the low achievers. More evidence is needed to support the Output Factor.

5. Input. Stronger support is needed to confirm teacher expectations relating to the Input Factor.

6. Touch. There was no evidence to show that teachers touched their high-achieving students more than they touched their low-achieving students. The frequencies of interactions were minimal and more evidence must be accumulated to provide additional support for the Touch Factor.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

The purpose of this study was to identify specific and differential teacher behaviors that affect student behavior based on Rosenthal's Four Factor Theory (1974). Specific purposes included the following:

1. The identification of teachers' differential treatment of students according to the teachers' expectations (high or low).

2. The identification of students' different responses according to the teachers' expectations (high or low).

3. The addition of the factor, Touch, as another possible identifying mechanism responsible for the Pygmalion Effect.

Teachers in four different physical education activity classes were asked to rank their students (total group in each class) in order of their physical achievement or skill potential. The rankings were used as the criterion measure of the teachers' expectations for their students' performance in physical education. Three observers using the Brophy and Good Interaction Analysis System (1969) recorded interactions on 96 (24 students from each class) junior high school students on six separate days within a two-week period. Forty-eight of the students were designated as high achievers and 48 of the students were designated as low achievers. Twenty-four frequency measures and 32 percentage

measures were derived from the coding. Analyses of Variance were performed on five variables and on types and quantity of contacts. The following results were obtained:

1. Quantity and types of contacts. A significant difference was found indicating that the designated high achievers were given more opportunities to respond and were asked more questions by the teachers.

2. Climate. A significant difference was found showing that teachers treated the designated high achievers more warmly than they treated the designated low achievers. Teachers discriminated between the two expectancy groups by displaying differential behaviors to the high and low achievers under equal situations.

3. Feedback. There was a significant difference in the amount of affirmation and praise indicating that teachers directed more evaluative comments to the designated high achievers. There was minimal evidence to show that teachers gave other kinds of feedback more to their designated high achievers than to their designated low achievers. The nonsignificant substatements outweighed the number of significant substatements. Because of this result, strong support could not be given for the Feedback Factor.

4. Output. A significant difference was evident indicating that the designated high achievers received more attention and were given more opportunities to respond. There was, however, no significant effect to show that teachers gave more reinforcement to the designated high achievers than they did to the designated low achievers. As a result, strong support could not be given for the Output Factor.

5. Input. There was no significant evidence to suggest that teachers taught more new material to the designated high achievers than they did to the designated low achievers.

6. Touch. There was no significant evidence to support the contention that teachers touched their designated high achievers more than their designated low achievers. Teachers did not exhibit any more Climate, Feedback, Output, or Input when they touched the designated high-achieving students than they did when they touched the designated low-achieving students.

Conclusions

The findings of the present study resulted in the following conclusions:

1. Teachers treat students differently according to the teachers' expectations. Total amount of attention, total dyadic contacts, and total procedural contacts indicating objective differences confirm Brophy and Good's (1969) findings.

2. Teachers who expect superior achievement from their students do treat them differently from their inferior achievers by exhibiting more warmth, praise, and support both in quantity and quality of interactions. These findings are supportive of other studies (Dalton, 1969; Kester, 1969; Page, 1970).

3. Teachers who expect superior achievement from their students treat them differently from their inferior achievers by exhibiting more feedback when appropriate responses are given. This one finding indicating differentiated feedback of

affirmation and praise concurs with Brophy and Good's (1969) results. Further study is indicated and more evidence is needed to strongly substantiate the Feedback Factor of teacher expectations.

4. Teachers who expect superior achievement from their students treat them differently from those from whom less is expected by exhibiting more Output. More attention and increased encouragement of responsiveness support conclusions found in studies by Brophy and Good (1969), Davis and Levine (1970), and Rubovitz and Maehr (1971). Further study is indicated for this factor for the purpose of determining the different types of reinforcement corresponding to particular responses.

5. Teachers who expect superior achievement from their students do not treat them differently from their inferior achievers by giving more Input. This finding was contrary to results reported by Beez (1970), Rist (1970), and Rosenthal (1974).

6. Teachers do not touch their designated high-achieving students more than their designated low-achieving students nor do teachers exhibit more Climate, Feedback, Output, and Input when they do touch their designated high-achieving students.

Implications

Rosenthal (1974) has stated that more research is needed ". . . to shed light on the mechanisms serving to mediate

interpersonal expectation effects" (p. 24). The contributions of the present study have increased further the probability that expectancy effects occur in an educational setting. Additional support is needed, however, to corroborate Rosenthal's Four Factor Theory (1974).

Replications of the present study are needed for two reasons. First, this study was the only study to have been conducted in physical education. More evidence is needed to substantiate the present findings. In addition, other factors that were not examined in the present study may need to be looked at to determine the extent of the operation of the Pygmalion Effect. The following suggestions that may be used in further studies are:

1. Extension of observation time. It may be possible to collect more information over a long period of time. If expectancy effects are examined two times per week, over an eight-week block of time, results may possibly be different. Changes of skill patterns and moods of students and teachers may be factors that emerge after a certain length of time. In addition, many of the differences between the high and low achievers for the majority of substatements which were not significant were in favor of the high achievers. It is possible that with more frequent interactions, the direction of the effect could become significant.

2. Measuring attention time. How long (in minutes) that a teacher interacts with a student may be a determinant of how students react to the teacher.

3. Separation of verbal and non-verbal climate. Different kinds of information may be obtained by observing students' responses to different kinds of climate. Students may react differently to nonverbal climate (winking, nodding, smiling) in contrast to verbal climate.

4. Selection of classes. Specific activity classes could be selected (after a specific period of time) to insure an optimum amount of interactions relating to selected variables.

5. Anecdotal record. Anecdotal evidence at the end of each week of observation time may add to the collected data.

6. Sex and race factors. Additional information may be collected by observing the teachers' differential behaviors according to the teachers' expectations relating to sex and race.

The second reason for replication is that the present study was conducted in a natural setting and no induced expectations were given to the teachers. In studies where expectancies are manipulated, the teachers are given information about their students that may not necessarily be true. The teachers are led to believe that one group is different from the other group. In these studies, the data have indicated support for the expectancy that was given to or held by the teacher. Studies are needed to determine if results of teacher expectations in natural settings produce the same kinds of differences or effects as when teacher expectations are experimentally manipulated.

BIBLIOGRAPHY

REFERENCES

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

Aiken, L. Review of pygmalion in the classroom. Educational and Psychological Measurement, 1969, 29, 226-228.

Allport, G. W. The role of expectancy. In H. Cantril (Ed.) Tensions that cause wars. Urbana, Illinois: University of Illinois Press, 1950, 43-78.

Alpert, G. Therapeutic effects of therapist-patient matching and positive therapist expectancies. Unpublished doctoral dissertation, Harvard University, 1970.

Baker, J. P., & Christ, J. L. Teacher expectancies: A review of the literature. In J. D. Elashoff & R. E. Snow (Eds.), Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971, 48-64.

Beez, W. V. Influence of biased psychological reports on teacher behavior and pupil performance. In M. W. Miles & W. W. Charters, Jr., (Eds.), Learning in social settings. Boston: Allyn & Bacon, Inc., 1970, 328-334.

Brophy, J. E., & Good, T. L. Teacher-child dyadic interaction: A manual for coding classroom behavior. Report series No. 27. University of Texas: Research and Development Center for Teacher Education, 1969.

Brophy, J. E., & Good, T. L. Teachers communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 1970, 61, 365-374.

Brown, W. E. The influence of student information on the formulation of teacher expectancy. Unpublished doctoral dissertation, Indiana University, 1969.

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

Carter, R. M. Locus of control and teacher expectancy as related to achievement of young school children. Unpublished doctoral dissertation, Indiana University, 1969.

Chaiken, A. L., Sigler, E., & Derlega, V. J. Nonverbal mediators of teacher expectancy effects. Unpublished doctoral dissertation, Old Dominion University, 1972.

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

Clark, K. B. Educational stimulation of racially disadvantaged children. In A. H. Passow (Ed.), Education in Depressed Areas. New York: Columbia Press, 1963, 142-162.

Coles, R. What can you expect? The New Yorker, April 9, 1969, 173-177.

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

Dalton, W. B. The relations between classroom interaction and teacher ratings of pupils: An exploration of one means by which a teacher may communicate her expectancies. Peabody papers in Human Development, 1969, 7, 6.

Davis, D. J., & Levine, G. The behavioral manifestations of teachers' expectations. Unpublished manuscript, Hebrew University of Jerusalem, 1970.

Elashoff, J. D., & Snow, R. E. Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.

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

Flanagan, J. C. Test of general ability: Technical report. Chicago: Science Research Associates, 1960, 78-99.

Fleming, E. S., & Anttonen, R. G. Teacher expectancy of my fair lady. American Educational Research Journal, 1971, 8, 241-252.

Flowers, C. E. Effects of an arbitrary accelerated group placement on the tested academic achievement of educationally disadvantaged students. In R. Rosenthal & L. Jacobson (Eds.), Pgymalion in the classroom. New York: Holt, Rinehart & Winston, 1968, 59-60.

Gage, R. In J. D. Elashoff & R. E. Snow, Pgymalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971. Foreword.

Gess, L. R. The effects of information which is provided to teachers concerning students on the attitudes and the behaviors of the teachers and the students. Unpublished doctoral dissertation, University of Michigan, 1969.

Goldberg, M. Factors affecting educational attainment in depressed urban areas. In A. H. Passow (Ed.), Education in depressed areas. New York: Columbia University Press, 1963.

Goldsmith, J. W., & Fry, E. The test of a high expectancy prediction on reading achievement and IQ of students in grade 10. In J. D. Elashoff & R. E. Snow, Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971, p. 51.

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

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

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Jose, J., & Cody, J. J. Teacher-pupil interaction as it relates to attempted changes in teacher expectancy of academic ability and achievement. American Educational Research Journal, 1971, 8, 39-49.

Kester, S. W. The communication of teacher expectations and their effects on the achievement and attitudes of secondary school pupils. Unpublished doctoral dissertation, University of Oklahoma, 1969.

King, A. S. Self-fulfilling prophecies in training the hard-core: Supervisors' expectations and the underprivileged workers' performance. Social Science Quarterly, September, 1971, 369-378.

Kohl, Herbert. Great expectations. The New York review of books. New York: NYREV, Inc., 1968, 31.

Lanzetta, K. T., & Hannah, J. E. Reinforcing behavior of 'naive' trainers. Journal of Personality and Social Psychology, 1969, 11, 245-252.

Leacock, E. B. Teaching and learning in city schools: A comparative study. New York: Basic Books, 1969.

Lunt, J. A procedure for systematically describing teacher-student verbal and nonverbal interaction in the teaching of choreography. Unpublished doctoral dissertation, University of North Carolina at Greensboro, 1974.

McClelland, D. C., & Winter, D. G. Motivating economic achievement. New York: Free Press, 1969, 349.

McLean, B. A. Unpublished data on teacher expectations in R. Rosenthal. On the social psychology of the self-fulfilling prophecy: Further evidence for pygmalion effects and their mediating mechanisms. New York: MSS Modular Publications, Inc., 1974, Module 53, 1-28.

Meichenbaum, D. H., Bowers, K. S., & Ross, R. R. A behavioral analysis of teacher expectancy effect. Journal of Personality and Social Psychology, 1969, 13, 306-310.

Merton, R. K. Social theory and social structure. Illinois: Free Press, 1957, 421-436.

Page, J. W. Experimenter-subject interaction in the verbal conditioning experiment. Unpublished doctoral dissertation, University of Toronto, 1970.

Palardy, J. M. What teachers believe-what children achieve. Elementary School Journal, 1969, 69, 370-374.

Passow, A. H. (Ed.). Education in depressed areas. New York: Columbia University Press, 1963.

Pfungst, O. Clever Hans (the horse of Mr. Van Osten): A contribution to experimental, animal, and human psychology. Translated by C. L. Rahn. New York: Holt, Rinehart and Winston, 1911, reissued in 1965.

Pitt, C. C. V. An experimental study of the effects of teachers' knowledge or incorrect knowledge of pupil IQs on teacher attitudes and practices and pupils' attitudes and achievement. Dissertation Abstracts, 1956, 16, 2387-2388. Unpublished doctoral dissertation, Columbia University, (was left out).

Plutchik, R. Foundations of experimental research. New York: Harper & Row, 1974.

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

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

Rosenthal, R. Empirical vs. decreed validation of clocks and tests. American Educational Research Journal, 1969, 6, 681-691. (a)

Rosenthal, R. Another view of pygmalion. Contemporary Psychology, 1970, 15, 524.

Rosenthal, R. Teacher expectations and pupil learning. In R. D. Strom (Ed.). Teachers and the learning process. Englewood Cliffs, New Jersey: Prentice-Hall, 1971, 33-60.

Rosenthal, R. Teacher expectations and their effects upon children. In G. S. Lesser (Ed.). Psychology and educational practice. Glenview, Illinois: Scott, Foresman, 1971, 64-87. (b)

Rosenthal, R. The pygmalion effect lives. Psychology Today, September, 1973, 57.

Rosenthal, R. On the social psychology of the self-fulfilling prophecy: Further evidence for pygmalion effects and their mediating mechanisms. New York: MSS Modular Publications, Inc., 1974, Module 53, 1-28.

Rosenthal, R. Correspondence to Patricia B. Crowe, July 9, 1975.

Rosenthal, R. Experimenter effects in behavioral research. New York: Irvington Publishers, Inc., 1976.

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

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

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom. New York: Holt, Rinehart & Winston, 1968.

Rosenthal, R., & 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., & Rosnow, R. L. Artifacts in Behavioral Research. New York: Academic Press, 1969, 181-277. (b)

Rosenthal, R., & Rubin, D. B. Pygmalion reaffirmed. In J. D. Elashoff & R. E. Snow (Eds.). Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971, 139-155.

Rothbart, M., Dalfen, S., & Barrett, R. Effects of teacher's expectancy on student-teacher interaction. Journal of Educational Psychology, 1971, 62, 49-54.

Rowe, M. B. Science, silence, and sanctions. Science and Children, 1969, 6, 11-13.

Rubovitz, P. E., & Maehr, M. L. Pygmalion analyzed: Toward an explanation of the Rosenthal-Jacobson findings. Journal of Personality and Social Psychology, 1971, 19, 197-203.

Safrit, M. Evaluation in physical education. Englewood Cliffs: Prentice-Hall, Inc., 1973, 125-126.

Schrank, W. The labelling effect of ability grouping. Journal of Educational Research, 1968, 62, 51-52.

Schrank, W. A further study of the labelling effect of ability grouping. Journal of Educational Research, 1970, 63, 358-360.

Seaver, W. B., Jr. Effects of naturally induced teacher expectancies on the academic performance of pupils in primary grades. Unpublished doctoral dissertation, Northwestern University, 1971.

Shore, A. L. Confirmation of expectancy and changes in teachers' evaluations of student behaviors. Dissertation Abstracts, 1969, 30, 1878-1879.

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

Snow, R. E. Unfinished pygmalion. Contemporary Psychology, 1969, 14, 197-200.

Spence, J., Underwood, B., Duncan, C., & Cotton, J. Elementary statistics. New York: Appleton-Century-Crofts, 1968.

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

Thorndike, R. L. But you have to know how to tell time. American Educational Research Journal, 1969, 6, 692.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Co., 1971.

APPENDIXES

APPENDIX A

Adaptation of Brophy and Good's Observation System

PREFACE

Brophy and Good (1969) have indicated that their observation should not ". . . be conceived as a finished, closed system to be used without modification" (p. 7). Research questions in the present study were relatively different from the questions posed by Brophy and Good in their classroom study. As a result, a different approach to coding some of the variables was required and additions, deletions, and extensions were made in some of the major divisions.

The following pages describe each category of the three main divisions with explanations for the adaptations of some of the categories. Symbols are used to identify the different types of changes. The symbol (*) indicates an extension of the category, and the symbol (**) indicates an addition to the category.

DIVISION I: RESPONSE OPPORTUNITIES

Response opportunities are defined in the system in three ways: (1) they are public interactions between the teacher and the student, but these interactions are intended for the whole class even though only one child is singled out for the interaction; (2) they occur when the teacher asks a question requiring a verbal or movement response from the student; and (3) only a single individual responds to the question. Response opportunities are teacher-afforded involving individual recognition by the teacher and they involve single questions that demand single responses (Brophy and Good, 1969).

Category I: Type of Response Opportunity

Type of response opportunity refers to the demand made upon the student. The teacher deliberately attempts to get a student to respond. The five types of response opportunities are Discipline, Direct, Open, Call Out, and Student-Initiated.

Subcategories

Discipline. Discipline questions are questions which the teacher uses to control behavior. A teacher calls on a student to force him or her to pay better attention.

Example: John, you will not do this correctly if you don't listen!

Direct. Direct questions are definite teacher-afforded response opportunities. The teacher calls on a student who has not indicated a willingness to respond by any overt action such

as raising a hand. The most obvious direct question is when the teacher calls out the student's name before asking the question.

Example: John, show me an air dribble.

Open. The teacher asks a question, waits for a show of hands and then calls on one of the students who has indicated a willingness to respond. The open question is partly afforded (teacher) and partly created (student) since the teacher asks the question and then waits for the student to raise his or her hand.

Example: Who can describe a lay-up shot to the class?

Call Out. Call out responses are opportunities created by students who do not wait for the teacher to call on them. The teacher asks a question and a student calls out an answer before the teacher gives the student permission to respond. The teacher must recognize the student who calls out the answer for it to be considered a call out response opportunity. The teacher must give some kind of feedback to that student when he or she calls out.

**Student-Initiated. This question is initiated by the student and does not involve the student answering a question posed by the teacher. The student may ask the teacher some additional information about the task involved or ask for some kind of help in a task.

Addition. Students at the secondary level speak out more than students at the elementary level. These students start to to become more active in the academic process by initiating actions.

Example: Miss Smith, am I doing this right?

Category II: Level of Question

Level of question refers to the nature of the response demand made upon the student. The five levels are identified as General Task, Process, Product, Choice, and Self-Reference. The first four refer to questions about academic or school-related content. The last question refers to the student's opinion or reaction not related to the content of the subject being discussed.

Subcategories

**General Task. This type of question or demand was added to the system for ease of coding and to minimize any confusion relating to Level of Question. This subcategory is comparable to any of the other four questions. Many times in physical education activities, the teacher explains a movement task to the whole class just as a classroom teacher may explain something in arithmetic or English. The physical education teacher then instructs the class to work on that particular task. Without posing any further questions, the teacher gives feedback to a performer. For ease in coding, the coder enters general task, then indicates whether the type of response opportunity was direct, open, or call out. The level of question is also indicated. It is easier for the judge to code general task when no obvious question is asked. A task demand is comparable to any level of question.

Example: Class, today I would like you to work on the lay-up shot. Please get in groups of

eight and start as soon as you pick up your equipment. I will move to each group to check on your progress.

Process. The student explains something that requires him or her to integrate facts or to move through a problem-solving process. These questions are "why" or "how" questions, which usually require extended sentences. They cannot be answered with a single word.

Example: Why is it important to bend your knees when picking up an object?

Product. Product questions only require knowledge of a specific fact. Usually the student answers with a single word or short phrase. These questions usually begin with: who, what, when, how much, how many, or where?

Example. Who is Chrissie Evert?

Choice. Choice questions involve "either-or" or "yes-no" questions or questions with more than two alternatives. The answer is always in one of the alternatives.

Example. Is it safer to pick up an object with the knees bent or with the knees straight?

Self-Reference. This question does not involve academic content as the choice questions must. Self-reference questions include opinions, preferences and personal information and do not require the student to give a correct factual answer.

Example: Do you like this activity?

Category III: Quality of Answer-Movement

The student answers a question in four particular ways: Correct, Incorrect, No response, and Don't Know. The teacher's intent is the criterion for what is correct or not correct.

Subcategories

Correct (+). If the student answers a question in such a way as to satisfy the teacher, the answer is correct. The teacher does not have to positively affirm an answer or make some favorable remark to the student's response. The answer is considered correct unless the teacher indicates some dissatisfaction either by giving the answer to the student or asking someone else to answer the question.

Incorrect (-). The answer is considered wrong if the teacher gives the answer to the student or asks another student to answer the question, or in any way implies dissatisfaction with the student's response. The teacher does not have to imply or tell the student that the answer is wrong.

No Response. The student does not respond when asked a question by the teacher. The student does not make any attempt to answer.

**Don't Know. The student implies that he does not know the answer to the question.

Addition. There is a distinction between letting the teacher know about the answer and not answering at all. The student who says that he or she does not know may be implying something different from the student who does not acknowledge the teacher's question.

Category IV: Atmosphere

Subcategories

**General Touch. A teacher moves toward a student and pats student on back, squeezes student's arm, puts arm around student's shoulder or touches any part of the student's body indicating warmth or a feeling of friendliness.

Addition. One of the teacher variables under study. This particular variable, Touch, was not in Brophy and Good's category system.

**Incidental Touch. The teacher touches the student while demonstrating a skill. Teacher stops class to explain a skill and may touch a student while talking.

Addition. One of the teacher variables under study. This particular variable, Touch, was not in the Brophy and Good category system.

Category V: Terminal Feedback

Terminal feedback implies feedback that is brought to a close. When a teacher gives a terminal feedback reaction to the student, he or she is either giving the answer to the student or making an evaluative response without giving an answer. The first six subcategories do not involve substantive responses. The last four subcategories do not have a substantive quality and provide some information to the student from either the teacher or from another student. More than one category in the terminal feedback section may be coded.

Subcategories

*Climate-Praise (‡). Climate-praise refers to the teacher's reaction which is more than affirmation or positive feedback. The teacher communicates a warm personal reaction and compliments the student by saying, "good", "wonderful" and "fine." The teacher verbally or nonverbally connotes a warm, friendly feeling to the student.

Adaptation. Climate is one of the teacher variables under study. The word, climate, was not in the system. The subcategory, praise, was considered comparable to the teacher variable, climate.

Example: That's very good, Mary!

Affirms Right. The teacher indicates that the student's response is correct by verbally saying "yes", "okay", "right." Nonverbal gesture of shaking head up and down is also coded as an affirm right. If the teacher repeats the student's answer, that is also considered a correct answer.

Example: Yes, that's right!

No Feedback Reaction (0). The teacher does not make any kind of response to the student's answer. The teacher does not indicate affirmation or negation verbally or nonverbally. A check mark is entered in the answer column as correct.

Negate Wrong. Negation parallels affirmation. The teacher either indicates verbally with the word "no" or "uh, uh", or "that's not right." A nonverbal gesture of shaking the head back and forth from side to side is also coded as negate wrong. The teacher is giving impersonal feedback concerning the correctness of the response.

Example: No. Susie, you're doing it wrong!

Criticism (=). Criticism goes beyond a simple "no" answer. The teacher expresses anger verbally or nonverbally. The teacher may indicate to the student that she is frustrated, hostile or disgusted.

Example: John, you can't do anything right! (accompanied by a look of anger or disgust)

**Assist Touch. The teacher moves to manually manipulate the student in a particular movement or touches a student to help with a task if the student is having difficulty.

Addition. One of the teacher variables under study. This particular variable, Touch, was not in the Brophy and Good category system.

Process Feedback. The teacher reviews the question with the student and explains how the student can arrive at the answer. The teacher does more than provide the student with the answer. Process feedback may follow wrong or right responses. Process feedback also follows a process question.

Example: If a process question such as "Why is it important to follow through after making contact in the forehand drive?" is not answered correctly, the teacher may give the answer but in addition go through the steps with the student to show her the effect of follow through on the speed of the object.

Gives Answer. The teacher gives the student the answer and does not elaborate as in the process feedback column. The "gives

answer" column is coded only when the student gives the wrong answer or has not answered the question or has not done the movement properly.

Example: John, do it this way.

Asks Other. The teacher does not give the answer but instead asks another student to help with the answer or movement.

Example: Mary, will you show Susie the correct grip?

Call Out. The call out column refers to a student who calls out the answer when the teacher asks another student for the answer. Before the questioned student can answer the question, another student calls out the answer.

Category VI: Sustained Feedback

The categories of sustaining feedback include the teacher behavior which extends and prolongs the response opportunity by giving the student a second chance. Sustained feedback indicates that the teacher, for whatever reason, prefers to stay with the student to help the student find an answer.

Subcategories

Repeat. The teacher repeats the question or asks the student to repeat the movement again. This takes place when the student looks perplexed or when the teacher wants to see the movement again. The answer or movement may be correct or incorrect.

Example: John, let me see that lunge again. I think your foot is too far forward.

*Rephrase Suggest-Correct. The teacher sustains the response by either giving a clue to the student to simplify the question,

or making a suggestion on how the student can move. The teacher may also correct a movement or part of a movement to help the student.

Adaptation. The suggestion-correction phase was added to this subcategory to allow for situations when the teacher helps the student in her moves by making small corrections or suggestions.

Example: That's good, Joanie! Stretch those arms just a bit more and it will be perfect!

New Question. The teacher asks a new question when an answer is required that is different from the original question. A question that requires a new answer is coded as a new question.

Example: Yes, Jane. Now show me how you can balance on another part of your body.

DIVISION II: PROCEDURAL CONTACTS

The category of procedural contacts includes all dyadic teacher-child interactions which involve permission to do something, supplies, equipment, reporting information to the teacher, getting particular information from the student, doing errands for the teacher or other kinds of classroom management.

Category I: Student-Created Procedural Contacts

The student is the initiator in the contact. Created contacts are planned by the student only and the teacher has not sought the child out for an interaction.

Subcategories

**Procedural Touch. The teacher moves toward the student to place the student in line or moves student to a particular spot on the floor.

Addition. One of the teacher variables under study. This particular variable, Touch, was not in Brophy and Good's category system.

Praise (+). The teacher communicates a warm personal reaction and compliments the student verbally.

Example: Thank you for moving that equipment. You did a fine job. (student offered to move equipment)

Feedback. The teacher responds in some way to the student's needs without praising or criticizing.

Example: Yes, you may move to the other end of the gym.

Criticism (-). No, you may not help Joan to take the equipment back. You're too careless.

Category II: Teacher-Afforded Procedural Contacts

The teacher is the initiator in the contact. Afforded contacts by the teacher usually have to do with classroom management. The teacher seeks out a student and asks the student to aid in getting out equipment, supplies, taking roll for the class or going on some kind of errand.

Subcategories

**Procedural Touch. The teacher moves toward the student to place the student in line or moves student to a particular spot on the floor.

Addition. One of the teacher variables under study. This particular variable, Touch, was not in Brophy and Good's category system.

Feedback. Teacher asks student for help. A check is entered in the feedback column.

Example: John, please help Jane take out the basketballs.

DIVISION III: BEHAVIORAL CONTACTS

Behavioral contacts refer to the student's behavior. Behavioral contacts are teacher-afforded and the interactions concern the student's behavior only.

Category I: Teacher-Afforded Contacts

Subcategories

Praise (+). Praise for behavior may sometimes occur. A student may be praised for being quiet.

Example: John, you were very good in class today. You were quiet and listened.

Warning. The student is singled out by the teacher. The teacher makes a comment about the student's inappropriate behavior.

Example: Jean, you're too noisy. You will not be able to hear the instructions.

Criticism (-). The teacher singles out a child and makes an angry or exasperated comment.

Example: John, I am not going to talk to you again. Sit Down!

**Behavior Touch. Teacher touches a student when giving the student a warning or criticism or praise of the student's behavior.

APPENDIX B

Mean Percentages and Analyses of Variance for Variables Input and Touch

Table 28

Analysis of Variance of Qualitative
Measures of Variable Input

Variable	Source	Df	SS	Ms	F
New questions following right answers over total right answers	Level	1	9.30	9.30	.06
	Class	3	806.89	268.96	1.60
	Level x Class	3	169.46	56.48	0.34
	Error	84	14150.06	168.45	
	Total	91	15135.73		

Table 29

Mean Percentages of Qualitative Measures for Variable Input
Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
New questions following right answers over total right answers	47	6.12	45	5.48	24	4.01	24	2.11	22	9.08	22	8.52

Table 30

Analyses of Variance of Quantitative Measures of the Variable Touch

Variable	Source	Df	SS	Ms	F
General touch	Level	1	1.80	1.80	1.11
	Class	3	16.33	3.36	3.36*
	Level x Class	3	10.59	2.18	2.18
	Error	88	142.59	1.62	
	Total	95	171.31		
Assist touch	Level	1	.19	.19	0.04
	Class	3	100.28	33.42	7.54**
	Level x Class	3	13.96	4.65	1.05
	Error	88	390.17	4.43	
	Total	95	504.60		
Procedure touch	Level	1	.04	.04	0.05
	Class	3	.11	.03	0.04
	Level x Class	3	7.52	2.50	2.82*
	Error	88	78.30	.88	
	Total	95	85.97		
Incidental touch	Level	1	.01	.01	0.96
	Class	3	.03	.01	1.01
	Level x Class	3	.03	.01	1.00
	Error	88	.91	.01	
	Total	95	.98		
Behavioral touch	Level	1	.04	.04	0.78
	Class	3	.12	.04	0.70
	Level x Class	3	.11	.03	0.65
	Error	88	5.33	.06	
	Total	95	5.60		

* .05 level
** .01 level

Table 31

Mean Frequencies of Quantitative Measures for Variable Touch Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total general touches	48	.56	48	.27	24	.12	24	1.12	24	.20	24	.20
Total assist touches	48	1.16	48	1.20	24	.58	24	2.87	24	.20	24	1.08
Total procedural touches	48	.47	48	.50	24	.45	24	.45	24	.54	24	.50
Total incidental touches	48	.02	48	.00	24	.00	24	.04	24	.00	24	.00
Total behavior touches	48	.04	48	.08	24	.08	24	.08	24	.00	24	.08

Table 32

Analyses of Variance of Qualitative Measures
for Variables Climate and Touch

Variable	Source	Df	SS	Ms	F
General touch following correct answers over total answers	Level	1	30.05	30.05	.83
	Class	3	1008.14	336.04	9.31**
	Level x Class	3	201.03	67.01	1.86
	Error	87	3141.77	36.11	
	Total	94	4381.99		
General touch following wrong answers over total answers	Level	1	.01	.01	.00
	Class	3	41.58	13.86	1.19
	Level x Class	3	57.69	19.23	1.66
	Error	87	1009.33	11.60	
	Total	94	1108.61		
General touch following negations over wrong answers	Level	1	1.05	1.05	.03
	Class	3	186.57	62.19	1.67
	Level x Class	3	31.11	10.37	.28
	Error	81	3020.75	37.29	
	Total	88	3239.48		

** .01 level

Table 33

Mean Percentages of Qualitative Measures for Variables Climate and Touch Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
General touch following right answers over total answers	48	3.29	47	2.16	24	.95	24	8.29	23	.28	24	1.28
General touch following wrong answers over total answers	48	.94	47	.92	24	.00	24	1.81	23	1.17	24	0.76
General touch following negations over wrong answers	45	1.11	44	.89	23	.00	21	3.57	22	.00	23	0.62

Table 34

Analyses of Variance of Qualitative Measures on Variables Feedback and Touch

Variable	Source	Df	SS	Ms	F
Assist touch following correct answers over total answers	Level	1	481.52	481.52	2.67
	Class	3	2008.04	667.01	3.72**
	Level x Class	3	461.40	153.80	.85
	Error	87	15671.60	180.13	
	Total	94	18622.56		
Assist following wrong answers over total answers	Level	1	35.74	35.74	.14
	Class	3	7551.59	2517.19	9.64**
	Level x Class	3	158.39	52.79	0.20
	Error	87	22728.19	261.24	
	Total	94	30473.91		
Assist touch following negation over wrong answers	Level	1	1.99	1.99	.04
	Class	3	705.22	235.07	4.40**
	Level x Class	3	177.32	59.10	1.11
	Error	81	4326.95	53.41	
	Total	88	5211.48		
Assist touch over total incorrect answers	Level	1	291.76	291.76	.22
	Class	3	39421.54	13140.51	10.02**
	Level x Class	3	3528.82	1176.27	.90
	Error	81	106272.05	1312.00	
	Total	88	149514.17		
Assist touch following give answer over total incorrect answers	Level	1	1.04	1.04	.01
	Class	3	9577.57	3192.52	16.34**
	Level x Class	3	67.87	22.62	.02
	Error	81	15825.35	195.37	
	Total	88	25471.83		

** .01 level

Table 35

Mean Percentages of Qualitative Measures for Variables Feedback and Touch Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Assist touch following correct answers over total answers	48	2.23	47	6.74	24	.77	24	11.58	23	.00	24	5.31
Assist touch following wrong answers over total answers	48	7.62	47	8.84	24	2.96	24	23.40	23	1.31	24	4.94
Assist touch following negation over wrong answers	45	2.34	44	2.04	23	0.00	21	7.19	22	.69	23	1.24
Assist touch over total wrong answers	45	18.95	44	15.33	23	5.60	21	54.86	22	2.05	23	8.74
Assist touch following give answer over total wrong answers	45	6.86	44	6.64	23	1.82	21	25.39	22	.85	23	.31

Table 36

Analyses of Variance of Qualitative Measures
for Variables Output and Touch

Variable	Source	Df	SS	Ms	F
Rephrase and assist touch following right answers over total right answers	Level	1	1.29	1.29	.38
	Class	1	8.43	2.81	.82
	Level x Class	3	13.07	4.35	1.27
	Error	84	288.33	3.43	
	Total	91	311.12		
Failure followed with sustained feedback and assist touch over total failures	Level	1	93.61	93.61	.60
	Class	3	1486.39	495.46	3.19*
	Level x Class	3	27.81	9.27	.06
	Error	81	12589.44	155.42	
	Total	88	14197.25		

* .05 level

Table 37

Mean Percentages of Qualitative Measures for Variables Output and Touch Calculated According to Expectancy Group and Class

Variable	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Rephrase and assist touch following right answers over total right answers	47	.13	45	.37	24	.00	24	.26	22	.00	22	.75
Failure followed with sustained feedback and assist touch over total failures	45	4.74	44	6.79	23	3.78	21	12.32	22	1.20	23	6.08

Table 38

Analysis of Variance of Qualitative Measures for Variables Input and Touch

Variable	Source	Df	SS	Ms	F
New questions and assist touch following right answers over total right answers	Level	1	34.61	34.61	1.21
	Class	1	285.34	95.11	3.33*
	Level x Class	3	179.99	59.99	2.10
	Error	84	2399.00	28.55	
	Total	91	2898.94		

* .05 level

Table 39

Mean Percentages of Qualitative Measures for Variables Input and Touch Calculated According to Expectancy Group and Class

	High Group		Low Group		Class 1		Class 2		Class 3		Class 4	
Variable	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
New questions and assist touch following right answers over total right answers	47	1.09	45	2.31	24	.00	24	2.65	22	.00	22	4.17

APPENDIX C

Sample Practice Questions for Training Judges

PRODUCT QUESTIONS

Product questions seek to elicit a single correct answer which can be expressed in a single word or short phrase. They differ from process questions in that they only require knowledge of a specific fact. They do not force the student to integrate several facts or to make inferences from them. Product questions usually begin with:

WHO - is Chris Evert?

WHAT - is a bogey?

WHEN - do you change sides of court in tennis?

HOW MUCH - swing is needed to get to the pin?

HOW MANY - points in a tennis game?

WHERE - do you place your hands in the headstand?

PROCESS QUESTIONS

Process questions are "why" or "how" questions and usually require extended phrases or sentences. The student explains something in a way that requires him to integrate facts or to show knowledge of their interrelationships. Process questions require the student to explain at length the cognitive or behavioral processes to be gone through in solving a problem or producing the correct answer to a question. They cannot be answered with a single word or short phrase.

WHY - is it important to bend your knees when picking up an object?

HOW - would you go about teaching or helping your classmate to learn the two-step?

CHOICE QUESTIONS

Two criteria distinguish choice questions: (a) the question deals with academic content and cannot be classed as a self-reference question, (b) the teacher provides response alternatives, either verbally or by showing the child visual aids to look at in connection with the question which includes the correct answer among them. Included are yes-no questions, either-or questions, and questions which present more than two alternatives.

EITHER-OR - Would you stand this way or that way?

YES-NO - Is this grip the same as this, or Is my grip the same as hers?

SELF-REFERENCE-OPINION

Any question which does not involve academic content and/or does not intend to elicit a particular correct factual answer. This includes opinions, preferences, and personal information.

Do you like activity?

Do you like this class?

The following questions are not in any kind of order. I thought it might be helpful not to put them under product, process, etc. Look them over and write what kind of question they are and hopefully we'll all come up with the same answers. Just follow the examples on the first page and you won't have much trouble. I hope to have three films ready for you on Thursday.

1. How many serves are you allowed to take from each side of the court in tennis?

2. What is a let serve?

3. What is a foot fault?
4. What is a double fault?
5. How many games in a set?
6. What does it mean when I say "sole your club?"
7. Who brought tennis to the U. S.?
8. When was tennis introduced to the U.S.?
9. Who is the leading woman tennis player in the U.S.?
10. Who is Billie Jean King?
11. Where is the U. S. Open played?
12. From what court do you serve when the score is even?
13. What is a lob?
14. What is a lay up shot?
15. Show me a jump shot.
16. Name three different types of passes.
17. Show me the difference between a forward stride position and a side stride position.
18. Who was the originator of the game of basketball?
19. Where was the game of basketball introduced?
20. Where is the deltoid muscle?
21. What muscles are involved in the sit-up?
22. Why is it important to warm up before strenuous activity?
23. Why is the overhand throw recommended for distance?
24. How can you correct a slice in golf?
25. Why is it important to follow through in the golf swing?
26. Is it important to follow through in the tennis swing?
27. What should you do if you lean to one side in the forward roll?

28. What is one problem if you cannot balance in the headstand?

29. How many pins must you set up for bowling?

30. Why is it more advantageous to knock the seven pen down from the right side of the lane?

31. Do you like Bowling?

32. What is a strike?

33. Why is it important to practice foul shooting in basketball?

34. Why is it important to give with the impact when receiving an oncoming object?

35. Why is it more difficult to weave and dribble than to dribble straight onward?

36. How does the action of the flat serve differ from the action of the slice serve?

37. What is the maximum possible score in bowling?

38. Why is the four-step approach in bowling preferable to the three- or the five-step approach?

39. Show me how you would balance on two body parts?

40. Is it easier to balance on more or less body parts?

41. Why is it easier to balance on two body parts than one body part?

42. When I say, "Keep your eye in the ball," do I really mean "Keep your eye on the ball?

43. Why is it important to know about angle of rebound in shooting?

44. Is it important to know about follow-through in all activities?

45. WHY?

46. Explain why it is important to keep your head down when putting.

47. What is the power phase of a stroke?

48. Why is it important to keep your arm out of the water on the recovery phase of the front crawl?

49. Does Newton's third law make sense to you in swimming?

50. WHY?

51. Apply Newton's third law in the elementary back stroke. Legs only.

52. If you did not glide at all in the side stroke, how would you cut down on the efficiency of the stroke?

53. Give me one example of inertia.

54. Show me the difference between the elementary back stroke and the inverted breast stroke.

55. Which is easier - the butterfly or the breast stroke?

56. WHY?

57. What is more important - style or efficiency?

58. WHY?

59. If you push down on the water what happens to your body?

60. If you are a thin person, can you keep yourself floating more efficiently?

61. HOW?

62. Why do you need to keep your wrist firm as you make contact in the forehand?

63. Why must you keep your eye on the ball until after contact?

64. What other move can you do before you land on the bench?

65. What else can you do on the ladder?

66. Why do you have your grip like that?

67. Who taught you that grip?

68. Where should the thumb be on the club?

69. What kind of stance should you use in the pitch shot?

70. Do you think that you can jump at least 10 feet?

71. Is that a safe move?

72. What do you need to correct to make this move a little more refined?

73. How do you hold the racket? This way or that way?

74. Would you stand this way or this way?

75. What is the proper stance to use when pushing a heavy object? Forward stride or side stride?

76. Are you more stable if you widen your base of support?

77. Define gravity.

78. Do you use an underarm or an overarm pitch in softball?

79. Is a bogey one over or one under "par?"

80. Is a five iron a pitching iron or a long iron?

81. Do you think dancing is fun?

82. Is the mazurka a relatively new step?

83. What is the two-step?

84. Is square dance the same as folk dance?

85. What is the difference between folk dance and square dance?

86. Show me the waltz.

87. Is it important to know how to dance?

88. Is it easier to pick up a heavy object with the knees bent or the knees straight?

89. Show me the correct way to pick up a suitcase.

90. Why is it important to bend your knees when picking up any kind of object?

91. Do you like to swim?

92. Why is it important to use the glide in the elementary back stroke?

93. Why do some people have shorter glides than others in the resting strokes?

94. Show me what happens when you push down on the water.

95. Now tell me what this means.

96. When do you breathe in the elementary stroke?

97. What is the easiest stroke for you? (This is not self-reference when referring to academic content. The teacher might second this question with WHY?)

98. Is it easier to float when you are thin or when you have more adipose tissue?

99. What is heavier? Fat or bone?

100. In the Elementary back stroke, why is it easier to move through the water when you keep your arms under on the recovery?

Some Thoughts

If a student is asked a product question such as: "Show me what happens when you push down on the water with your hand," and then is asked either "Why?", or "Tell me what this means", or "Explain that please," then the second coding would be process. So it would be product question first, then maybe some kind of feedback, and then process check, and probably some kind of feedback. This would be a continuous coding for that same person.

Other Types of Questions

What would you do if?

Is it safer to do _____ or _____?

What is your opinion on _____.

Is it easier to move this way or _____.

Show me two different ways to _____.

Is it possible to _____? How?

APPENDIX D

Sample Coding and Summary Sheets

CODING SHEET

1. Class _____ 2. Date _____ 3. Start ____ 4. ____ 5. Activity _____ 6. Attendance _____

7. Observer _____ 8. Page _____ of _____.

Response Opportunities																																										
Type of Question					Level of Question					Ans-Movt				Atmos		Terminal Feedback									Systm			S-Prc				Tprc		Behavior								
DISC	DIRECT	OPEN	CALL	STUDINIT	GEN TASK	PROCESS	PRODUCT	CHOICE	SELF REF	+ CORRECT	- WRONG	NO RESP	DON'T KNOW	GEN TOUCH	INC TOUCH	++ CLIMATE	AFF RIGHT	NO FDBCK	NEG WRONG	= CRIT	ASST TOUCH	PROCESS	GIVE ANS	ASK OTHER	CALL	REPEAT	REPH SUGG	NEW INFO	PROC TOUCH	+	FB	-	TOUCH	FB	+	W	-	TOUCH				
	17						✓			✓				✓	✓		✓																								
	19						✓				✓								✓				✓																		
	21						✓			✓						✓	✓										✓														
							✓			✓						✓	✓				✓						✓														
							✓									✓	✓																								
																																	✓	20								
																																				10					

Expectations:

Remarks:

SUMMARY SHEET

RESPONSE OPPORTUNITIES AND DYADIC CONTACTS

Student _____ Class _____

Obs.	Type Response					Question					Created Procedure				Afforded Procedure		Behavior			
	Stud. Init.	Disc.	Direct	Open	Choice	Gen. Tsk	Process	Product	Choice	Self reference	Touch	Feedback	+praise	-crit.	Touch	Feedback	+praise	-crit.	Warning	Touch
1.																				
2.																				
3.																				
4.																				
5.																				
6.																				
7.																				
8.																				

SUMMARY SHEET

TEACHER FEEDBACK IN RESPONSE OPPORTUNITIES

Student _____

1. + _____
2. - _____
3. nr _____
4. dk _____

Class _____

General Class Response Opportunity

Obs.	Atmos.		Terminal										Sustained		
	Gen Tch	Inc Tch	Cli-mate ++	Affirm Right	No Feed-back	Negate Wrong	Criti-cism =	Asst Tch	Pcss	Give Ans	Ask Oth	Call Out	Rpt	Reph Sugg	New Q Inform
1.															
2.															
3.															
4.															
5.															
6.															
7.															
8.															
9.															
10.															
11.															
12.															
13.															
14.															
15.															
16.															
17.															
18.															

EXPLANATION OF ENTRIES ON SAMPLE CODING SHEET

Activity: Tennis

Interaction One

Teacher: John, can you show me the forehand grip?

John: John demonstrates the forehand grip.

Teacher: Yes, John, very good! (Teacher puts arm around John's shoulder.)

Entry

John's code number is 17.

Enter number 17 in Direct column (teacher asked John to demonstrate). The next check mark is placed in the Product column (fact). Before a check can be placed in the Ans.-Movt. column, the coder must wait for the teacher's feedback. A check mark is entered in the Correct column (answer is right). Another check is entered in the General Touch column and Climate column and then the Affirmative Right column.

The interaction ends with Terminal Feedback.

Interaction Two

Teacher: Jane, when the score is even, do you serve from the left-hand court or the right-hand court?

Jane: I would serve from the left-hand court.

Teacher: No! The right-hand court.

Entry

Jane's code number is 19.

Enter number 19 in the Direct column, a check in the Product column, a check in the Wrong Answer column, a check in the

Negation column, and a final check in the Give Answer column.

The interaction ends with Terminal Feedback.

Interaction Three

The teacher has given a general task for all students to work on their tennis serve. A general task is comparable to a question.

Teacher: Betsy, that serve is very good! Now, toss the ball a little bit higher and let's see if you can make some aces.

Student performs:

Teacher: Yes! That's excellent. Extend that arm just a bit more.

Student performs:

Teacher: Very, very good, Betsy. You're improving every day.

Interaction ends.

Entry

Betsy's code number is 21.

Number 21 is entered in the Direct column followed by a check in the Product column. By affirming that the tennis serve is good, the coder enters a check mark in the Correct Answer column, followed by a check in the Affirmation and Climate columns. After the teacher praises the student, she continues to teach the student and give sustained feedback instead of ending the interaction. The teacher makes a suggestion to the student to toss the ball higher. A check mark is entered in the Rephrase column. The coder then drops down to the next row and enters a check mark in

the Product column. The coder does not enter the student's identification number again. By checking the Question column, the interpreter knows that this interaction was an extended one and not just one response opportunity. After the teacher makes the comment, "Yes, that's excellent!", the coder continues on the second row and makes check marks in the Correct Answer column Affirmation column, and the Climate column. The teacher makes another suggestion and says, "Now extend your arm a little bit more," and at the same time, the teacher helps to extend the student's arm. The coder now checks Assist Touch and Rephrase, and again drops down to the row below and enters a check in the Product column. When the teacher says, "Very good!", to the student, the coder enters a mark in the Correct Answer column, the Affirmation column, and the Climate column.

Interaction Four

The coder enters number 20 in the Teacher Procedure column. The teacher has approached the student and has asked him to perform some procedural task. The teacher touches the student as he talked to him, and a check mark is entered in the Procedural Touch column.

Interaction Five

The coder enters the number 10 in the Behavior Warning column. The teacher has given a warning to the student concerning his disruptive behavior.

APPENDIX E

Frequency and Percentage Measures Taken on Five Variables

FREQUENCY MEASURES

1. Direct questions.
2. Open questions.
3. Call out questions.
4. Student initiated questions.
5. Total response opportunities (sum of above, Questions 1-4).
6. Afforded procedure contacts (teacher).
7. Created procedure contacts (student).
8. Total dyadic procedure contacts (sum of above, Questions 6-7).
9. Behavioral praise.
10. Behavioral warning.
11. Behavioral criticism.
12. Total behavioral contacts (sum of above, Questions 9-11).
13. Total dyadic contacts (response opportunities, plus created procedures, plus afforded procedures, plus behavioral contacts. Sums of Questions 5, 8, and 12).
14. General touches.
15. Incidental touches.
16. Assist touches.
17. Procedure touches (afforded and created).
18. Behavioral touches
19. Total touches (sum of above, Questions 14-18).
20. Climate (warmth-praise).
21. Sustaining feedback (sums of repeat, rephrase, and new information).
22. Total correct answers
23. Total incorrect answers (includes no response and don't know).

24. Total answers (sum of above, Questions 22-23).

PERCENTAGE MEASURES

A. Measures of Teacher versus Student Initiation of Dyadic Interactions

1. Direct questions over response opportunities.

2. Open questions over response opportunities.

3. Call out questions over response opportunities.

4. Student initiated response opportunities over total response opportunities.

5. Created (student) procedure contacts over total procedure contacts.

6. Afforded (teacher) procedure contacts over total procedure contacts.

B. Level of Question Measures

1. Process questions over total questions.

2. Product questions over total questions.

C. Student Performance Measures

1. Correct answers over total answers.

2. Wrong answers, no response, and don't know over total answers.

D. Climate (Warmth-Praise) and Criticism of Academic Performance

1. Climate (warmth-praise) following correct answers over total answers.

2. Affirmation and climate (warmth-praise) of right answers over total right answers.

3. Climate following wrong answers (includes don't know and no response) over total answers.

4. Negations (including criticism) following wrong answers over total wrong answers.

5. Criticism following right answers over total answers.

6. General touch following correct answers over total answers.

7. General touch following wrong answers over total answers.

8. General touch following negation over wrong answers.

9. Assist touch following correct answers over total answers.

10. Assist touch following wrong answers over total answers.

11. Assist touch following negation over wrong answers.

E. Quality of Feedback (Terminal)

1. Number of no feedback over total correct responses.

2. Assist touch over total incorrect responses.

3. Assist touch following give answer over total incorrect responses.

4. Give answer over total incorrect responses.

F. Sustained Feedback After the Initial Response Opportunity

1. New questions following right answers over total right answers.

2. Rephrase following right answers over total right answers.

3. Repeat over repeat plus rephrase plus new question following failure (teacher demands response to original question rather than help student).

4. Failure followed with sustained feedback over total failures (teacher sticks with student in a failure situation rather than give answer).

5. Failure followed with sustained feedback and assist touch over total failures.

6. New questions and assist touch following right answers over total right answers.

7. Rephrase and assist touch following right answers over total right answers.