TEACHER BEHAVIORS ASSOCIATED WITH MATHEMATICS ACHIEVEMENT IN THIRD GRADE STUDENTS

by

Shirley Ann Cole

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 7 8

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my direction by Shirley Ann Cole

entitled Teacher Behaviors Associated with Mathematics Achievement in Third Grade Students

be accepted as fulfilling the dissertation requirement for the degree of Doctor of Philosophy

Ralph J. Weber Jr. — Dissertation Director — 7-21-78 — Date

As members of the Final Examination Committee, we certify that we have read this dissertation and agree that it may be presented for final defense.

Ralph J. Weber Jr	8/1/78
Sidney W. Bijou	8/1/78
Robert B. Bechtel	8/1/78
Tom Ratcliff	8/1/78
John R. Bergan	8/1/78

Final approval and acceptance of this dissertation is contingent on the candidate's adequate performance and defense thereof at the final oral examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Shirley Cole

ACKNOWLEDGMENTS

It is with a great deal of pleasure that I thank Dr. Ralph J. Wetzel for his contribution to this dissertation and to my graduate education. I also appreciate the support of Dr. Sidney Bijou, Dr. Thomas Kratochwill, Dr. John Bergan, and Dr. Robert Bechtel.

I would also like to thank Esther Milne of the Tucson School District #1 Research Office for her efforts in selection and recruitment of teachers. Finally, I much appreciate the participation of the eight teachers.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABSTRACT

Research on teacher characteristics which influence student achievement has been largely based on trait theories of teaching and learning. In general, this research has not been able to identify relations between teacher traits and student performance. The present study was guided by behavioral theory and attempted to determine if relationships between teacher behaviors and student achievement in the natural environment could be discovered.

Subjects were eight third grade teachers and the children in their classes. Teachers were grouped into four pairs, with socio-economic level of the school and performance of the students on the Metropolitan Reading Test as matching variables. Within each pair, one teacher had students who achieved higher scores on the Stanford Mathematics Test than the other teacher.

Between eight and twelve half-hour observations were made in each classroom during the mathematics period. University student observers used a ten-second interval format to record a variety of teacher and student behaviors, including three kinds of consequences (Positive, Acknowledge, and Negative), seven kinds of instruction (General Instruction, Probe, Describe Materials, Rule, Model, Repeat Instruction, and Give Correct Answer), and seven kinds of student responses (On-Task, Off-Task, No-Task-Available, Correct Responses, Wrong

Responses, Initiates-to-Teacher, and Questions). Comparisons in rates of these behaviors were made using t-tests.

With regard to consequent behaviors, teachers in the Higher Group showed a higher rate of Positive ($P < .05$), a higher rate of Positive combined with Acknowledge ($P < .005$), and a lower rate of Negative ($P < .05$). For each of these behaviors, the finding was true for at least three of the four teacher pairs. With regard to instructional behaviors, three of four Higher Group teachers had a higher rate of Rule, and a lower rate of General Instruction. These relationships approached significance ($P < .2$). Among student behaviors, two differences approached significance. These were Questions and Initiates-to-Teacher. Three of four Higher Group teachers had higher rates of these behaviors in their classrooms.

These results support the contention that relationships between teacher behaviors and student achievement can be identified in the natural environment. The case was strongest that higher rates of praise result in higher achievement. Comparisons are made between the rates of Positive and Negative which were recorded in this study and those of a previous study which were also recorded in the natural environment. The rates recorded in this study for Lower Group teachers were very similar to those reported in the previous study. Higher Group teachers had a much higher rate of Positive.

In addition, a comparison of three methods of presenting data on inter-observer agreement was made. Two agreement statistics are presented: percent on scored intervals and Phi coefficient. Also the data of the calibrating observer were presented.

The comparison showed these results: the average correlation between percent on scored intervals and Phi was .96. The correlations between rates recorded by primary observer and the calibrating observers on days of agreement checks ranged between .78 and .98 ($\overline{X}$ = 92). T-tests on calibrator data were highly similar to tests on primary observer data for agreement check days.

It was concluded that Phi and percent on scored intervals provide very similar information. The problems of demonstrating agreement for very low rate behaviors are discussed.

CHAPTER 1

INTRODUCTION

It is commonly held in our society that teachers make a significant contribution to the quality of education in our schools. For example, higher pay for teachers is a way to attract more qualified personnel to the field. Presumably, more qualified teachers will produce greater achievement in students. Poor student performance in poverty areas is, in part, attributed to the fact that low pay and crowded classes do not attract good teachers. The tenure system is criticized for making it more difficult to replace low quality teachers. Finally, there is criticism of the university that faculty are hired and promoted for their research efforts rather than teaching ability, and that the quality of instruction thereby suffers. In all of these the assumption is that the teacher has a major impact on student achievement.

Trait Theory Approach

This cultural belief is reflected in most psychological theories about teaching, and a substantial amount of research has been directed at identifying relations between teacher characteristics and student achievement (Heath and Nielson, 1974; Rosenshine and Furst, 1971). Most of these studies have developed out of trait theories of psychology.

The theories assume that global descriptions of people can be discovered which will predict their behavior in diverse settings, and which will be stable over time. These traits are internal personal characteristics and control overt behavior. Behavior change is effected by changing the underlying characteristics.

Researchers using this theoretical basis have investigated such variables as clarity, variability, enthusiasm, task-oriented and/or business-like behaviors, student opportunity to learn criterion material, use of student ideas and general indirectness, criticism, use of structuring comments, types of questions, probing, and level of difficulty (Rosenshine and Furst, 1971). Definitions of these variables have been determined by the underlying theory, i.e., global and not derived directly from observations of the behavior of the teacher. For example, Solomon, Bezdek, and Rosenberg (1963) scored a teacher as variable depending on the answers to two questions: "Did the instructur follow up new and unexpected topics, suggested by student questions or suggestions?" and "Did he (she) sometimes make changes in the way material was being presented?" Also, in accordance with the theory, the variables have been measured by rating scales (Belgard, Rosenshine and Gage, 1968), adjective checklists (Fortune, 1967), Q-sorts (Wallen, 1966), supervisor ratings, self-ratings and direct observation (Medley and Mitzel, 1959). When direct observations are made, the categories are typically combined into scales which include a number of diverse behaviors. For example, Medley and Mitzel (1959) derived numerical scores for verbal emphasis (high score indicating class engaged in reading and writing tasks), and

social organization (high score indicating little teacher talk and class organized in several groups).

This kind of research is exemplified in a study by Cunningham (1975). He studied over a thousand kindergarten students and their 108 teachers. Students were typified along eight measures including two ability tests, three rating scales and three demographic variables. Teachers were typified along five variables, including one rating scale and four demographic variables. Four student and four teacher types were extracted and each subject was then classified as belonging to one type. Student types included, for example, extroverted/Black/female, and slow/alienated/male; teacher types included Black/experienced, and inexperienced/student-centered/empathic. Analysis of variance was then used to determine if teachers of a given type were equally effective in raising achievement with all types of children. He found a significant main effect for student types and a significant interaction between student and teacher types. The author concluded that homogeneous groupings of students with appropriately selected teachers might improve student achievement.

Overall, this sort of research has produced very inconclusive results. Both Rosenshine and Furst (1971) and Heath and Nielson (1974) conclude that teacher characteristics which are associated with student achievement have not been identified. Heath and Nielson go so far as to suggest that the relationship between student demographic and sociological characteristics and achievement would be a more fruitful area of research.

This model has also been used in an attempt to identify counselor traits associated with client improvement in therapy and student aptitude-treatment interactions which lead to improved achievement in special education. Similar inconclusive results are reported in these areas (Rowe, Murphy and DeCsipkes, 1975; Ysseldyke, 1973).

Behavior Theory Approach

One of the predominant theoretical alternatives to trait theory is behavior theory. This theory also predicts that a teacher could have a major effect on student achievement. The theory assumes that behavior varies as a function of antecedents and consequences in the environment which surrounds the behavior. Research in this model focuses on behavior defined in objective terms and relies on direct observation of frequency, duration, and intensity rather than judgments by observers in the form of ratings.

A behavior analysis suggests three areas in which a teacher has the opportunity to improve student achievement. These include antecedents to student behavior, consequences to behavior and generalization of new learning.

Antecedents include organization and sequencing of materials based on an analysis of the skill, establishment of prerequisite behaviors and the direct instruction to the child. To profit from a learning experience, a child must previously exhibit a certain number of skills, both academic and social. For example, before a child can learn to add, he must be able to count. Before learning from a lecture, a child must attend to the teacher. It is also important that each component

of a new skill is taught before it is required in the solution of the problem. That is, multiplication must be taught before division because a child must multiply in order to complete a long division problem.

The direct instruction of the child includes several teacher behaviors: cues or instructions, models or demonstrations, and props or physical guidance. A good teacher would move from using props to cues as rapidly as possible while maintaining a high rate of correct responding.

The second area in which teachers might affect student achievement is behavioral consequences. It is assumed that children will repeat behaviors which are followed by positive reinforcers and not those which are followed by neutral or punishing consequences. An efficient learning environment will be one in which correct academic responses and the supporting social behaviors are strengthened and incorrect responses which distract from learning would be consequated so as to decrease in frequency. Teachers can affect consequences both by the management of their own social behavior and by the arrangement of materials and activities.

Finally, the responses of the children must generalize beyond the teaching situation. Teachers can structure the learning environment so that children are more likely to exhibit their new skill in the environment where it will be most useful.

The present research focused on two areas of teacher behavior: direct instruction and behavioral consequences. Specifically, it attempted to identify relations in the natural environment between

teacher cues, models and other instructional behaviors, reinforcers and both the achievement of the students and achievement-related student behaviors.

Although the importance of teacher behaviors has not yet been demonstrated in the natural environment, a great deal of experimental research has shown their importance in the laboratory and experimental-field situations.

Consequences

Researchers have demonstrated experimentally that teacher consequences can affect both achievement-related behaviors and achievement. Hall, Lund and Jackson (1968) showed that teacher attention contingent on on-task behaviors increased those behaviors. Long and Williams (1973) and Packard (1970) demonstrated that class related tangible rewards, such as free time, could also increase on-task behavior when presented contingently. Schmidt and Ulrich (1969), Barrish, Saunder and Wolf (1969) and Harris and Sherman (1973) showed that disruptive behavior could be reduced by the same reinforcers contingent on low rates of disruptive behavior.

Other studies demonstrate that contingent reinforcers can improve performance directly. Hughes (1973) studied seventh graders in three specially constructed science lessons. He found that social consequences from the teacher for answering questions (both praise and mild reprimands) significantly improved performance on a posttest. Hops and Cobb (1974) showed improved reading scores when praise was contingent on volunteering and on-task behavior. Walker and Hops (1976)

demonstrated that greater achievement was produced when reinforcement was contingent on similar achievement-related behaviors or academic progress. Finally, Chadwick and Day (1971) showed improved California Achievement Test scores when teacher attention and tokens were contingent on appropriate social behavior and increased speed and accuracy on class assignments.

These studies provide convincing evidence that achievement (as measured by standardized achievement tests) can be improved by appropriate contingent use of reinforcement.

Instruction

The research on instructional variables is considerably less extensive. Most research investigates the influence of an entire teaching approach or of a reinforcing technique rather than antecedent instructional variables (Rowbury, Baer and Baer, 1976).

A number of authors (Harris and Sherman, 1973; Long and Williams, 1973; Medland and Stachnik, 1972; Packard, 1970) investigated the influence of instructions versus reinforcement. They uniformly found that instructions without contingent reinforcement had no long lasting effect on behavior. Herman and Tramontana (1971), however, demonstrated that instructions could greatly enhance the effectiveness of reinforcers. They provided token reinforcers to children for good nap time behavior without previous instruction about how tokens would be dispensed. This produced slow improvement in appropriate resting. The addition of instructions resulted in a large and immediate increase in good napping.

This suggests that instructions can contribute to the effect of contingent reinforcement.

Although most research focuses on treatment packages, some effective instructional behaviors have been identified. Of these, the best documented is modeling. Bandura (1969) and his associates have demonstrated changes in a variety of behaviors in clinical settings due to modeling. Studies by Brigham and Sherman (1968) and Hart and Risley (1968) showed increases in the targeted classes of language behavior in preschool children when modeling was combined with reinforcement. Lahey (1971) demonstrated large increases in descriptive adjectives after modeling without explicit reinforcement.

Another effective instructional variable is prompting. Knapczyk and Livingston (1974) instructed two retarded students to raise their hands and ask questions when they did not understand an instruction or a word in a reading assignment. This resulted in more question asking, more on-task behavior and improved reading comprehension. Glynn and Thomas (1974) found more on-task time when the teacher displayed a poster indicating what behaviors were considered on-task at any given time (either seat work or attending to lecture). This prompt was an addition to an on-going self-observation and reinforcement program. Hardiman, Geotz, Reuter and LeBlanc (1975) produced increases in motor activity in a handicapped preschool child with the use of suggestions that she engage in those activities. Finally, Broden, Copeland, Beasley and Hall (1977) found that teacher use of questions which required multi-word answers increased the length of student responses.

Student Responses

Research has demonstrated that certain student classroom behaviors are related to achievement, both in the natural environment and in experimental classes. Two studies, using multiple regression techniques, have isolated student behaviors in nonexperimental classes. Cobb (1972) studied mathematics and verbal achievement as measured by the Stanford Achievement Test in two fourth grade classes. Important predictors of achievement included attention-to-task, task-relevant talk to peers, self-stimulation and compliance.

Soli and Devine (1976) used the same observation system with 312 high and low achieving third and fourth graders. They predicted verbal achievement on a subsample of the Gates MacGinitie and math achievement on a nonstandardized test devised for the study. Important predictors included task-relevant talk to peers, self-stimulation, failure to attend to task, initiation to teacher, inappropriate locale and play. The cumulative correlation using the four most powerful predictors for all children was .41.

Hops and Cobb (1974), as reported above, went on to demonstrate that increases in these behaviors would result in improved achievement.

Summary and Statement of Problem

Psychological theory and widespread cultural belief concur that teachers should make an important contribution to the achievement of their students. Research based on trait theories of teaching, however, has not been able to confirm this belief. Research based on behavior theory provides experimental support for this proposition. However,

principles which have been demonstrated effective in the laboratory have not necessarily been adopted or practiced in the typical public classroom. As of yet, there has not been a demonstration of the relation between teacher behavior and achievement in the natural environment of public schools.

This study attempted to provide this demonstration in two areas: behavior consequences for academic behaviors and direct instructions. Three types of teacher social reinforcement were included as forms of consequences: praise, neutral acknowledgment and reprimand (including feedback to children that responses were incorrect). Praise and reprimand were included because the literature has shown them to be powerful variables in controlling behavior. Acknowledgment was included because pilot observations indicated that it was a high frequency category and that it usually followed correct responses. Thus, if these fairly neutral statements function as positive reinforcers, they would have a powerful effect because of their frequency and contingent relationship to student responses.

Seven categories of instruction were included: modeling, rules, description of materials, repetition of instructions, telling a child the correct answer; probes and general instructions. Modeling was included because the literature has demonstrated that it is an effective teaching device. Rules, description of materials and repetition of instructions were included as forms of prompting which might characterize teacher behavior. Rules included teacher instructions which would allow a student to determine whether a response was correct. This category

was of particular interest because it would allow a student to get feedback on a response independent of the teacher. Probes were responses by the teacher which required students to answer. Good sequencing of materials requires that teachers have frequent feedback on each child's performance in order to make individual adjustments in the sequence. This category was intended to measure the amount of information the teacher evoked about student performance. General instructions included general lecture behavior which was not recorded under another category.

In addition, seven student behaviors were investigated in this study. These were included because the research literature indicates that improved achievement related to teacher behavior is often accompanied by changes in achievement-related student behavior. The student behaviors studied here were: On-Task, Off-Task, No-Task-Available, Initiates to Teacher, Questions, Correct and Wrong responses. Several of these were included because they have been demonstrated to be important in previous research (On-Task, Off-Task, Initiates to Teacher). Correct and Wrong responses were included based on the behavior analysis of student behavior. Theory suggests that student responding is an important requirement in the acquisition of new skills. It provides teachers with opportunities to strengthen correct responses and discover areas of difficulty (Skinner, 1968).

Observer Agreement

A second concern of this study was the method of calculating and presenting data on inter-observer agreement. It is generally agreed that high inter-observer agreement is one way to guard against

the problems associated with human observers. Although not infallible, high agreement helps insure that observers record the events as they occur. Low agreement suggests that the data may be distorted by inaccurate recording, observer bias and drift. There has been considerable controversy over the mathematical calculation which best represents the correspondence between two sets of data (Hartmann, 1977; Kratchowill and Wetzel, 1977).

For many years the most common calculation was what is now called "percent agreement interval by interval" (Hawkins and Dotson, 1975). This is calculated by counting as agreements each cell in which the two observers make identical entries, either both or neither recording the behavior; this number is divided by the total number of cells and converted to a percent.

This method has been criticized (Hawkins and Dotson, 1975; Johnson and Bolstad, 1973) for being potentially misleading, especially for low rate behaviors. It is possible for two observers to never agree on the occurrence of a behavior and yet achieve a high percent agreement interval by interval if they otherwise agree on its nonoccurrence. For example, suppose two observers record for ten intervals with the first observer recording the behavior in the first cell, the second recording in the last cell, and neither recording in the intervening eight. The agreement would be 80%. This is a respectable, though not a high figure, and might be interpreted as adequate evidence that two observers had recorded reasonably similar phenomena. Critics believe that is an unwarranted assumption, and the method problematic.

In response to this criticism, two other calculations have been suggested (Hawkins and Dotson, 1975). The first is called "percent agreement on scored-intervals"; the second, "percent agreement on unscored intervals." These are calculated by leaving out a certain number of intervals in the denominator--either the cells in which neither observer recorded (for percent on scored-intervals) or in which both observers recorded (for percent on unscored intervals). Thus, the figures are not inflated if the behavior is very low or very high frequency. This procedure does, of course, make agreement harder to achieve.

There are some difficulties with both these methods. First, a complete picture cannot be achieved since they exclude some data from the calculation. Second, it is mathematically easier to get a high percent on scored intervals if the rate of the behavior is either very high or, for unscored intervals, very low, and more difficult if the behavior is of a medium rate. This is because the opportunity for chance agreements increases as the rate increases. Thus, if two observers each record instances of a behavior in nine out of ten intervals, they will necessarily agree eight times. A percent on scored intervals of 80% for this base rate does little to inspire confidence in the data.

On the other hand, agreement of scored intervals is very difficult to achieve if the behavior is very low rate. At these rates a few disagreements have a large effect on the percent. For example, when the primary observer records one instance of a behavior, and the

calibrator records none, agreement on scored intervals is zero. A change of only one interval on the part of the calibrator would result in agreement of 100%.

Several suggestions have been made in light of these problems. Some (Gelfand and Hartmann, 1975; Kratochwill and Wetzel, 1977) have suggested using Phi, which is a correlation-like statistic that makes use of all the data and accounts for chance agreement. A Phi of 1.0 indicates perfect agreement; a Phi of .00 indicates only chance agreement; and a Phi of -1.0 indicates perfect disagreement. Kratochwill and Wetzel (1977) have also suggested presenting the data from both observers. They note that agreement statistics are intended to allow the reader to judge the similarity between the two sets of data without having to refer to them. If the agreement statistic does not adequately serve this purpose, it may be necessary to return to the primary data. They suggest that presenting the calibrator data allows the reader to determine by visual inspection whether the primary observer differed systematically from the calibrator in such a way as to call into question the conclusions of the study.

Attempts have been made to follow these suggestions in this study. In the Results, three figures will be presented: percent agreement on scored intervals, Phi, and the correlation between rates recorded by the primary observer and the calibrating observer on days of agreement checks. Tests of significance were calculated on primary observer data because they represented a larger sample of behavior.

CHAPTER 2

METHOD

Subjects

Subjects for this study were eight third grade teachers from a large urban public school system in Tucson, Arizona, and the students in their classes. Seven teachers were women, one a man. Teachers were selected as four matched pairs based on the Stanford Mathematics Test, Primary III Level. Within each pair, one teacher had students who obtained higher mathematics achievement scores than the other. Pairs were matched for correspondence on two variables--socioeconomic status of the school, as defined by the school district, and reading subtest score on the Metropolitan Achievement Test. The achievement data were collected in October of 1976; students tested were in the fourth grade. Thus children providing the achievement data were, in general, in the classes of the teachers the year before the testing. Because testing was completed early in the school year, it is assumed that the teachers primarily influencing the scores would be the previous year's third grade teacher rather than the present fourth grade teacher. Therefore, the first or High Group of teachers is one whose students are matched for socioeconomic level and reading ability, and who score higher in mathematics than do students of the second or Low Group.

Teachers were identified as potential subjects and recruited by the public school district research office. Teachers received a written explanation of the study which described the study as an exploratory effort to identify behaviors which might, in future research, serve to differentiate teachers who produce higher achievement in their students (Appendix A). They were not informed of the actual basis for selection of subjects or the categories of behaviors which would be recorded. A first round recruiting effort included written correspondence and some personal contacts between the research office and teachers. This resulted in four teachers willing to participate. A second effort included personal contacts by the experimenter, and produced the other four subjects. No major incentives, such as credit for inservice training, were offered for participation.

Procedure

Between eight and twelve observations were made in each class during the regularly scheduled mathematics period. Observations of about 30 minutes were made once or twice weekly and each included at least one five-minute observation focused on the teacher (Teacher Form) and one five-minute observation focused on the students (Child Form). These were repeated, in that order, if time allowed. Observers did not interact with teachers or students during the observations. They were introduced to the students as visitors who would be observing the class. The final eight observations in each classroom were included in data analysis.

Observers

Observers were six university students. Each had about a month of training before observations began. Three of the observers recorded in three different classrooms, two recorded in two classrooms, and one recorded in one classroom. Four of the six did agreement checks with two other observers; two with only one.

Observers received the same explanation of the study as did teachers. Specifically, they were not told that selection of teachers was based on achievement scores.

Data Collection

Data were collected using a ten-second interval and time sampling procedure. Ten teacher behaviors and seven child behaviors were recorded. These are listed in Table 1. A complete list of definitions for these behaviors is shown in Appendix B.

Data were collected using two formats, one which focused on the teacher (Teacher Form) and one which focused on the students (Child Form). The Teacher Form involved recording all teacher behaviors. Child behaviors were recorded only when a child interacted with the teacher. Thus, behaviors of children not involved directly with the teacher were not recorded. Child Form involved recording all the behaviors of a given child. Teacher behaviors were recorded only when the teacher interacted with the child under observation. Children were observed for three ten-second intervals at a time, in random order. (Four child behaviors were recorded only on Child Form, including Initiates to Teacher, On-Task, Off-Task, and No-Task-Available.)

Table 1. Teacher and student behaviors

	Description
Teacher Behaviors	
Probe	Most questions and commands
General	General instructions
Model	Demonstrations
Gives Rule	Statement of "how" or ways to determine if answer is correct
Describes Materials	Describes an object
Repeats Instruction	Gives an instruction again
Gives Correct Answer	Answers questions, or tells the correct answer
Positive	Clearly indicates an answer is correct
Negative	Indicates an answer is wrong
Acknowledges	Indicates hearing an answer
Child Behaviors	
Correct	Gives a correct response
Wrong	Gives an incorrect response
Question	Asks a question
Initiates to Teacher*	Initiates an interaction with the teacher
On-Task*°	Doing required task
Off-Task*°	Not working on required task
No-Task-Available*°	No academic task is assigned

*Recorded only on Child Form
*°Recorded with a time sampling procedure

The data collection format involved recording sequences of teacher antecedent behaviors, child responses, and teacher consequences. The categories represent a mutually exclusive and exhaustive set of categories. All the behaviors of the teacher and students were recorded. A particular category could be recorded more than once if another behavior intervened. Thus, a teacher might ask two questions (Probe) within one interval and they would both be recorded. A sample cell might look like this:

Teacher: What is this? (Probe)

Student: A bar graph. (Correct)

Teacher: Good, why is that? (Positive and Probe)

Interval 1
P P
C
+

If a behavior began in one interval and carried over into the subsequent interval, it was recorded in both intervals, thus:

Interval 1

Teacher: Today we are going to review . . . .

Interval 2

Teacher: . . . . subtraction (General)

Student: What page? (Question)

Teacher: Ninety-one. (Gives Correct Answer)

Interval 1	Interval 2
G	G
	Q
	GC

In most intervals no more than one teacher-child interaction occurred. A complete description of the procedure and sample data collection forms are shown in Appendix B.

Inter-Observer Agreement

Agreement checks were made, as allowed by the schedules of the observers. The number of checks for each class is shown in Table 2.

Table 2. Number of agreement checks in each class

	1	2	3	4	5	6	7	8
Number of Visits	10	12	10	9	10	10	9	12
Number of Checks	4	7	4	8	4	4	0	0

Two agreement figures were calculated for each behavior. These included Phi percent agreement on scored intervals. Data were collapsed into a two-by-two table as follows:

		Observer 2	
		Did Not Record	Recorded
Observer 1	Recorded	A	B
	Did Not Record	C	D

Each cell was scored as one of these four possibilities.

Agreement was calculated according to these formulas (Gelfand and Hartmann, 1975, pp. 211-212):

$$\text{Phi} = \frac{BC - AD}{\sqrt{(A + B)\ (C + D)\ (A + C)\ (B + D)}}$$

$$\text{Percent on Scored Intervals} = \frac{B}{A + B + C}$$

The correlation between rates for primary and calibrating observers for agreement check days is also presented.

CHAPTER 3

RESULTS

Data were analyzed using t-tests for matched pairs with three degrees of freedom. One-tailed tests were used when previous research suggested the direction of the relationship between the behavior and achievement. Two-tailed tests were used when no such prediction could be made. Differences between groups at the .05 level were considered significant. For one-tailed tests, differences at the .10 level are reported as trends. For two-tailed tests, differences significant at the .20 level are reported as trends. These lower level differences are reported because the t-tests in this situation are not powerful tests. This is due to the small sample size and to moderate differences between groups in achievement scores. This lack of power means that only a large difference between groups will be significant at conventionally accepted levels. Substantial confidence, therefore, can be placed in tests significant at .05 or better. Tests which are significant at .1 or .2 should be indications of areas for further research. The rates which are presented were calculated by dividing the frequency of the behavior by the total number of intervals and multiplying by 100.

Matching Variables

Achievement Scores

Teachers were selected for inclusion in this study based on the scores of their former students on the Stanford Mathematics Test and the Metropolitan Reading Test. Pairs were formed matching teachers for the reading, so that one teacher in each pair was higher on the mathematics score. The differences between the scores for each pair of teachers are presented in Table 3.

Table 3. Difference between achievement test scores between pairs of teachers in higher and lower groups

Pair	Metropolitan Reading		Stanford Math	
	1975	1976	1976	1977
1	-9	-6	+14	+2
2	-9	-6	+14	+2
3	-7	+6	+6	+8
4	-8	-10	+10	+6

These differences were calculated by subtracting the scores of the Lower Group teacher from the scores of the Higher Group teachers. Thus, a positive number indicates that the Higher Group teacher in the pair had a higher score. A negative number indicates that the Higher Group teacher had a lower score.

In general, the data indicate that teachers in the Higher Group had students who scored higher in mathematics ($\overline{X}$ = 9, range 2-14) than did teachers in the Lower Group. The reading scores indicate that teachers in the Higher Group scored lower ($\overline{X}$ = 6.1, range -10-+6). These data confirm that teachers in the Higher Group did have students performing better in mathematics. They also suggest that this higher performance is not a reflection of higher overall performance as indicated by their lower reading scores.

Socioeconomic Status

Teachers were also matched for socioeconomic status of the school. This was done because of the school district's belief, supported by Heath and Nielson (1974), that achievement is related to sociological variables such as income and subculture. Each school was assigned a rating by the district based on 14 demographic variables extracted from the 1970 census. For two pairs the Higher Group teachers were from a slightly lower SES level and for two pairs from a slightly higher one. Two teachers came from upper middle class schools; five came from middle class schools (including one at the upper end of the middle and one at the lower end); and one teacher came from a lower middle class school. These data indicate that SES level was independent of group membership.

Other demographic characteristics of the teachers are presented in Table 4. Teacher groups did not differ in age, educational level, years of teaching experience, or class size.

Table 4. Demographic characteristics

Variable	Higher Group Mean	Lower Group Mean
Age	45 years	43 years
Education	Master's plus 15 units	Master's plus 15 units
Experience	13 years	14 years
Class Size	23.7	26

Group Differences

Consequences or Reinforcement

Three of the teacher behaviors recorded were considered forms of reinforcement by teachers for the quality of student responses. These included Positive (praise and other enthusiastic indications that a student response was correct), Acknowledge (relatively neutral indications that the teacher had heard the response, generally meaning that it was correct), and Negative (reprimand and indications that a response was wrong). Based on previous literature, it was predicted that Higher Group teachers would use a higher rate of Positive and Acknowledge, and a lower rate of Negative.

In the Positive category, three of four teachers in the Higher Group had a higher rate. The mean difference between the groups was significant beyond the .05 level ($t = 2.346$). Phi for this category

averaged .66 (range 0-100), percent on scored intervals averaged 58% (range 0-100). The correlation between rates of the primary and calibrating observers on days of agreement checks was .89.

For Acknowledge, three of four higher teachers made more such statements, but not significantly so ($t = 1.29$, $P < .25$). Phi averaged .33 (range 0-100), percent on scored intervals averaged 28% (range 0-100) and the correlation between observers was .78.

When combined, Positive and Acknowledge show a stronger relation. All Higher Group teachers made more of these statements and the mean difference was significant at the .005 level ($t = 7.25$). These data are shown in Figure 1.

Negative showed the opposite relationship. Three of the Lower Group teachers had a higher rate of Negative. The mean difference was significant at the .05 level ($t = 2.47$). These data are presented in Figure 2. The mean Phi was .52, and the mean percent on scored intervals was 43% (range 0-100). The correlation between observers was .94. The rates for these three categories are presented in Table 5.

Direct Instructions

Seven behaviors were considered forms of instruction. These included General Instruction, Probe (most questions and commands to respond), Rule, Model, Describe Materials (telling about an object), Repeat Instruction, Give Correct Answer (answers questions, provides answer). No predictions were made about the relationships between these behaviors and achievement.

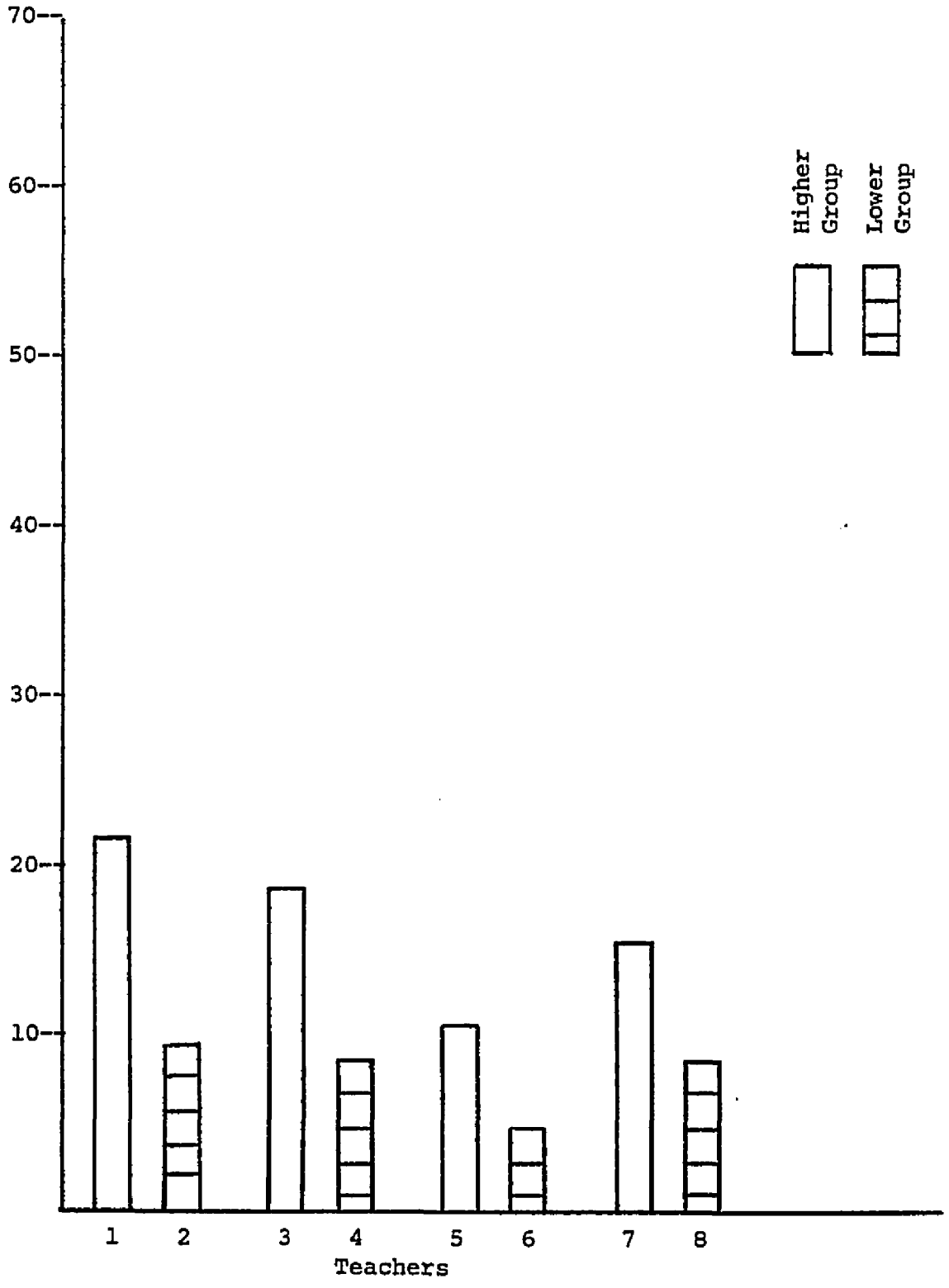

Figure 1. Percent of intervals scored for positive and acknowledge forms of reinforcement

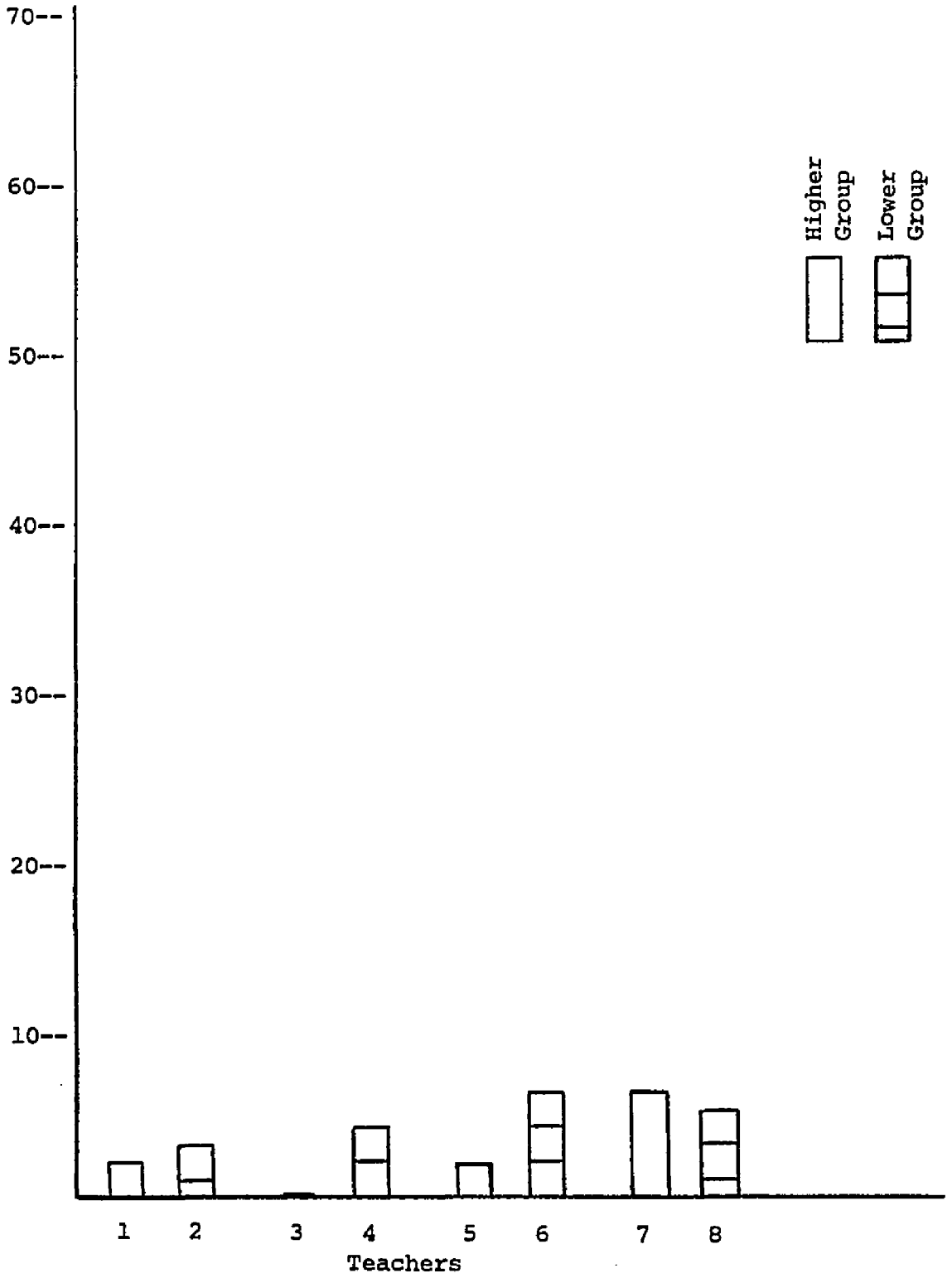

Figure 2. Percent of intervals scored for negative form of reinforcement

Table 5. Percent of intervals scored for positive, acknowledge, and negative

Pair	Positive		Acknowledge		Negative	
	Higher Group	Lower Group	Higher Group	Lower Group	Higher Group	Lower Group
1	12.6	02.7	09.5	07.5	01.6	03.3
2	15.0	03.7	03.0	05.4	00.2	04.1
3	03.7	03.5	07.1	01.2	02.4	05.6
4	08.6	05.4	07.5	04.1	05.5	05.4

Overall, there were no significant differences between groups in any of these behaviors. Only two approached significance (at the .2 level): General Instruction and Rule. For General Instruction, three of four Higher Group teachers had lower rates, and for Rule, three had higher rates.

There was an additional suggestion which might be of interest in future research concerning General Instruction. For the first two pairs of teachers, the pairs were assigned randomly. This was done because two teachers participated from each of two schools. For this reason, these two pairs could be rematched without disrupting the control for reading achievement or SES level. When this is done, significant differences emerge. All four Higher Group teachers show lower rates of General Instruction, and the mean difference is significant at the .01 level ($t = 6.25$). See Figure 3. When General Instruction

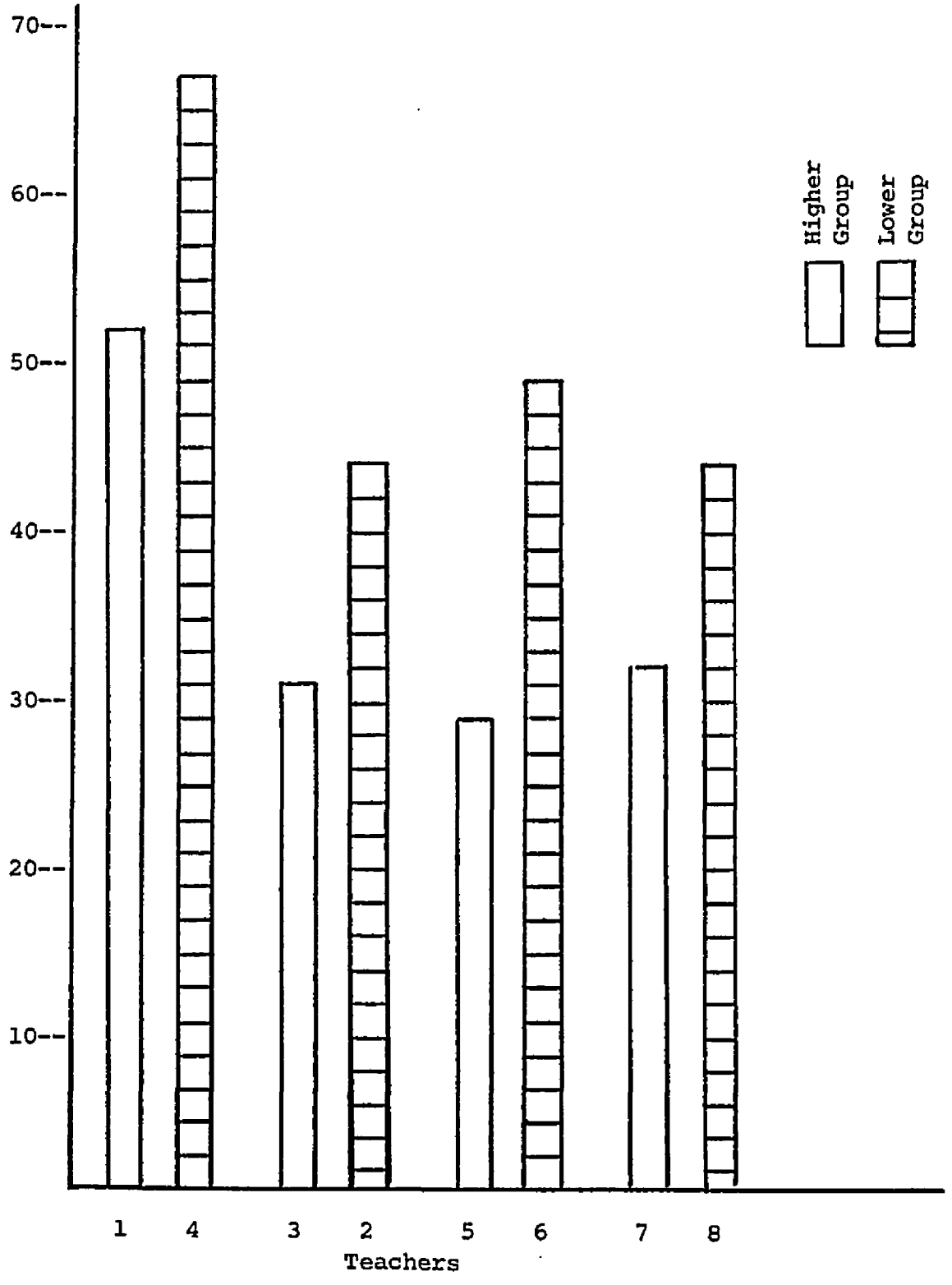

Figure 3. Percent of intervals scored for general instruction

is combined with Probe, the mean group difference is significant at the .1 level, and all of the Higher Group teachers have a lower rate. On the original match this combination approaches significance (P = .2).

In addition, when the remaining five categories of instruction are combined, all Higher Group teachers have a higher rate (t - 2.78, P <.1). See Figure 4. Rates for General Instruction and Combined Instructions are presented in Table 6.

The mean Phi was General Instruction was .75 (range -.09--1.00); the mean percent was 75% (range 15-100); and the correlation between observers was .88.

Student Responses

A total of seven student responses was recorded: Questions, Correct Responses, Wrong Responses, On-Task, Off-Task, No-Task-Available, and Initiates-to-Teacher (usually volunteers to respond). The first three were recorded on Teacher Form (i.e., they involved students as they interacted with the teacher). The last four were recorded on Child Form, and included a random selection of students. It was predicted that Higher Group teachers would show a higher rate of Questions, Correct Responses, On-Task, and Initiates-to-Teacher, and lower rates of Off-Task and No-Task-Available. No prediction was made for Wrong Responses.

Three behaviors approached significance. For Wrong Responses (t - 1.875, P <.2), Questions (t = 1.74, P <.1), and Initiates-to-Teacher (t = 2.057, P <.1), higher rates were recorded in three of

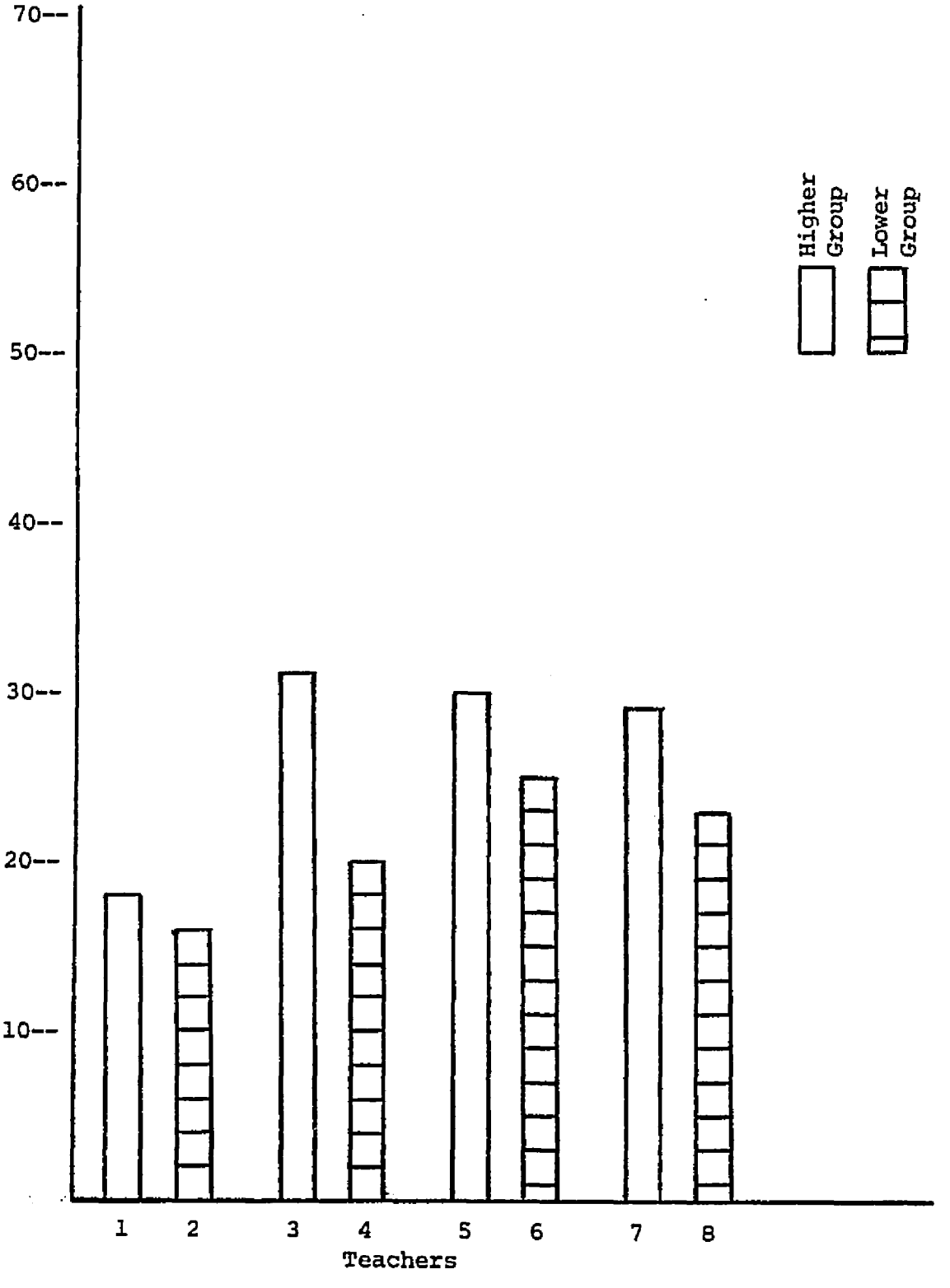

Figure 4. Percent of intervals scored for instructional categories of rule, model, describe materials, repeat instruction, and give correct answer

Table 6. Percent of intervals scored for general instruction and combined instructions

Pair	General		Combined	
	Higher Group	Lower Group	Higher Group	Lower Group
1	50.4	42.5	17.6	16.2
2	30.2	65.8	30.9	20.2
3	26.8	48.3	28.8	24.3
4	31.3	42.7	28.5	23.1

four Higher Group classes. Rates of these behaviors are reported in Table 7.

The average Phi for Wrong was .61, for Initiates was .77, and for Questions was .66. The average percent on scored intervals for Wrong was 53%, for Initiates was 77%, and for Questions was 55%. The correlation between observers for Wrong was .93, for Initiates was .95, and for Questions was .95.

Consent to Participate

A total of 37 teachers were approached as potential participants. Ten agreed to participate. A breakdown of this number by group membership is presented in Table 8. Fourteen potential Higher Group

Table 7. Percent of intervals scored for wrong response, questions, and initiates-to-teacher

Pair	Wrong Response		Questions		Initiates	
	Higher Group	Lower Group	Higher Group	Lower Group	Higher Group	Lower Group
1	04.0	03.9	10.2	04.7	01.6	00.9
2	02.8	01.8	03.3	01.0	04.3	00.3
3	02.2	02.2	15.3	02.9	0	0
4	05.2	03.9	00.2	00.6	08.3	05.1

Table 8. Consent to participate

Agreement to Participate	Group Membership	
	Higher	Lower
Yes	6	4
No	8	19

teachers were approached. This represents 38% of the total. Six or 43% agreed to participate. Twenty-three Lower Group teachers were approached. This represents 62% of the total. Four or 17% agreed. A chi square test for significance between these figures was significant at the .1 level (chi square = 2.873, df = 1).

CHAPTER 4

DISCUSSION

Group Differences

A large body of research has failed to confirm a widespread and pervasive belief in our society that teachers are an important influence on the achievement of children in school. Heath and Nielson (1974), after reviewing this literature, conclude that the effort to find relationships between teacher characteristics and student achievement has failed. This research, in general, developed out of trait psychological approaches. Research guided by behavioral theories of teaching has been more successful. Various studies cited in the Introduction provide experimental evidence that changes in teacher reinforcement and instructions can improve achievement and strengthen achievement-related behaviors. Up until now, however, there has not been a demonstration that teacher behaviors are related to achievement in nonexperimental situations, in the natural environment. The present study begins to provide that demonstration. It also provides information about what the important teacher behaviors are in improving achievement.

Reinforcement

The strongest finding in the present research was that teachers of higher achieving children use more positive social reinforcement.

This confirms the relevance to the nonexperimental classroom of a great deal of experimental research on the importance of positive reinforcement in controlling behavior. This finding supports the presumption that successful teachers use the reinforcement techniques that have been demonstrated effective, although they may not explicitly identify their procedures as positive reinforcement.

There is less of a research base relevant to the finding on aversive reinforcement, particularly in field situations. This is perhaps due to a general reluctance, based on ethical considerations, to experimentally introduce aversive stimuli in a classroom. The less frequent use of negative reinforcers by Higher Group teachers is supported by a finding reported by White (1975). He reported rates of approval and disapproval recorded in 104 classrooms, including ten third grade classes. For each grade he found that disapproval was a more frequent response than approval. In this study, that is only true of teachers in the Lower Group. For the Lower Group, three of four teachers had higher rates of the category Negative than of Positive. The fourth teacher had equal rates. In the Higher Group, all four teachers had lower rates of the category Negative than of Positive. Overall, the rate of Positive was approximately equal to the rate reported by White, but the rate of Negative was lower.

Antecedents

There has been less research on effective antecedent instructions than on reinforcement variables. Often the instructions given by teachers are studied in the process of demonstrating that the most

powerful controlling variable in the treatment package is the contingent reinforcement rather than the instructions. Other studies have shown the effectiveness of instructional techniques, such as explicit timing, which are not common in the public schools in the United States. This study provides some suggestions about methods of instruction which are more effective, and some information about teaching behaviors which do not appear important in relation to achievement.

The data indicate one positive finding about antecedents--that Higher Group teachers use less General Instruction. This category was included in the data collection system as a catch-all category, and included instructional behaviors which were not recorded under a more refined category, such as Model. It is likely that the higher achievement associated with less reliance on General Instruction resulted from those teachers using another more effective technique, and not from the reduced General Instruction directly. What this alternative might be is not indicated by the data. There were no significant differences between groups on any of the alternatives to General Instruction.

Of the instructional behaviors studied here, modeling has received the most research attention. Its effectiveness as a means of changing behavior is well documented. Research has been extended to investigate the parameters which contribute to its effectiveness (Bandura, 1969). This literature led to the prediction that a high rate of Modeling would distinguish the Higher Group teachers. The data did not support this prediction. One Higher Group teacher had a higher rate of Modeling, and two had lower rates than the Lower Group

teachers. One pair had the same rate. In general, the rate of Modeling was low for all teachers. The average rate across all teachers was 3.2%, the third lowest rate of all the teacher behaviors recorded. Three teachers had no recorded instances of Modeling. Clearly, this is not a widely used technique in these classes. Most of the teachers agreed with this assessment when shown the data. They indicated that they tend to ask children to respond in situations which might be appropriate for teacher modeling. It is possible that they rely on peer modeling rather than teacher modeling. There was no assessment of peer modeling in this study. The data do not provide a good test of the relationship between modeling and achievement in the natural environment. They rather suggest that modeling is an area of experimental research that has had little impact on teaching techniques in the classes studied.

Student Responses

Another belief about teaching that appears to be strong in this society is that school is a place where children ask questions. Question asking is seen as a natural result of interest and curiosity about a topic, and so question asking should characterize classes which are interesting to children (Sarason, 1971). The research literature, however, suggests that school classrooms are characterized by teacher questioning and student answering. Suskind (1969) reported student question asking as occurring at the rate of approximately two per half hour. The rate recorded in this study averaged considerably higher than that (mean

of approximately nine per half hour). The data indicate that while student questioning is a low frequency behavior (occurs less than 5% of the time) it may be associated with greater achievement on the part of the students. This trend toward significance may indicate a fruitful area for further research.

The present study was an attempt to demonstrate that teacher behaviors are associated with student achievement in the natural environment. Research on the association between student behaviors and achievement has a longer and more successful history. Research by Cobb, Hops, and associates (Cobb, 1972; Hops and Cobb, 1974) at the Center at Oregon for Research in the Behavioral Education for the Handicapped has identified a variety of behaviors that predict achievement in school. These behaviors include volunteering, attending, looking around, complying and relevant talk to peers. Three behaviors similar to these were investigated in the present study: Initiates-to-Teacher (volunteering), On-Task (attending), and Off-Task (looking around). Only the relationship between volunteering and achievement was supported by this study. The lack of a relationship between On-Task and achievement in this study may be related to the fact that On-Task was a high rate behavior in all the classes. The rates recorded here ranged between 70% and 90%. These are comparable to rates reported in the literature at the termination of treatment programs designed to increase on-task behavior.

In this study, one other behavior was related (though only as a trend) to achievement. Wrong Responses occurred more frequently in

classes of Lower Group teachers. This would be an intriguing finding, especially in light of the inverse relationship between reprimand and achievement; however, analysis of inter-observer agreement presented below calls this finding seriously into question. While the primary observer showed the reported result, the calibrator showed a stronger opposite effect.

Observer Agreement

Low Rate Behaviors

As was mentioned above, agreement is more difficult to achieve when the behavior is low rate. The agreement reported here undoubtedly reflects that fact. The percents on scored intervals are generally lower than those reported in journals (mean across all behaviors of 63%). The Phi's reach of surpass .60, the level recommended by Gelfand and Hartmann (1975) and Kratochwill and Wetzel (1977). The influence of the low rates of most of the behaviors is supported by the correlation of .60 between the rates of the behaviors and the percent on scored intervals.

These data bring up the issue of the usefulness of these figures as judgmental aids in determining the dependability of the data. The percents on scored intervals suggest that data are not sufficiently dependable. Hawkins and Dotson (1975), however, maintain that current standards were developed for a more lenient method and should be revised. The Phi's suggest that the data are dependable. But, the research community has little experience with Phi and so standards are

based on the recommendations of a few researchers, rather than experience with the statistic. The present situation appears to be one in which the primary data should be examined rather than relying on a judgmental aid.

This can be done in two ways in this study: first, by calculating the correlation between the rate of the primary and calibrating observers on days of agreement checks. A correlation would indicate the similarity in the daily picture presented by the two observers, regardless of the point by point agreement. Second, a new set of t-tests can be calculated on the calibrator data. If these show similar results, then more confidence can be placed in the original data.

The correlations between rates of the two observers are shown in Table 9. They range between .78 and .99 and average .92. They suggest confidence in the data.

T-tests based calibrator data are presented in Table 10. Two sets are presented--tests based on calibrator data and tests based on primary observer data for days of agreement checks. It was anticipated that tests on this last set of data would differ from the original tests because they are based on a much smaller sample of behavior. Similarity between the new primary tests and the calibrator tests would support the reliability of the original data.

The primary observer on agreement days showed results weaker than the original but in the same direction in nearly every case. The calibrator showed the same results as the primary observer on agreement days, with two exceptions: the calibrator supported the difference in

Table 9. Correlations in rates between calibrating and primary observer on days of agreement checks

Behavior	Correlation
Positive	.89
Acknowledge	.78
Negative	.94
General Instruction	.88
Rule	.99
Wrong Responses	.93
Questions	.95
Initiates-to-Teacher	.95

Table 10. Tests of significance based on data of primary observer on agreement check days and on data of calibrator

Behavior	Original Data	Primary on Agreement Days	Calibrator
Positive	.05	.1	.1
Combined Positive and Acknowledge	.005	.002	.1
Negative	.05	.025	.025
General Instruction	(.2)	(.2)	(.2)
Rule	.2	.5	.5
Combined Instruction	.1	.5	.5
Wrong Responses	(.2)	.5	.01
Questions	.1	.025	.1
Initiates-to-Teacher	.1	.025	.025

( ) Indicates a negative relationship

the rates of combined Positive and Acknowledge less strongly, and showed a strong opposite effect for Wrong Responses. These results lend confidence to the basic findings but call seriously into question the relationship between rate of Wrong Responses and achievement.

The problem with base rate and demonstrating dependability of data that was encountered is one that should be considered by the research community. Agreement is harder to demonstrate for low and high rate behaviors using either Phi or percent. At present, there is no

allowance made in publication standards to account for this. The effect, therefore, may be to reinforce differentially by publication research on medium rate behaviors. It may also mean less research on behaviors which are difficult to detect and more on very obvious behaviors. These are not necessarily the most interesting or socially significant behaviors. This appears to be a collateral effect of the demand for high agreement rather than one discussed and accepted by the research community. This is an issue which should be considered in establishing means of demonstrating the reliability of data.

Comparison of Phi and Percent

Jack Michaels (1974) described statistics as "judgmental aids" in responding to data. He advocates their use when they increase the efficiency of a researcher in responding to his or her data. He argues against their use when they interfere with such responding. He points out that if a statistic is very complicated a researcher may need to spend a good deal of time learning to respond to the statistic. He argues that, often, the time spent learning to respond to statistics is not returned in increased efficiency.

If a statistic is complicated, there is also the danger that the researcher will misunderstand the aid and make judgments based on the aid which are not valid. Any statistic reduces the amount of information to which the researcher responds. It is important to know what kinds of information are eliminated in order to use the statistic appropriately.

The present study allows a comparison of Phi and percent on scored intervals. Phi is recommended by certain superior mathematical properties, but is newer and more difficult to respond to than percent. The evidence here indicates that they present very similar pictures of agreement.

The greater difficulty in responding to Phi comes from the fact that its value is affected by a larger number of variables. Both percent and Phi are affected by the rate of the behavior and the number of agreements. If the rate is high, a high percent on scored intervals is easier to achieve because a few disagreements are relatively less important. If the rate is low, it is harder to achieve. For Phi, if rate is high or low, agreement is harder to achieve.

Phi is also affected by the discrepancy between the two observers in number of disagreements. If one observer is responsible for many more of the disagreements than the other, good agreement is easier to obtain. For example, if one observer records five instances in agreement with the other observer, and ten additional instances, which the second observer does not record, Phi will be higher than if each recorded five instances the other had not.

The similarity in the information provided by Phi and percent is attested to by the correlations between them. These are presented in Table 11. They average .96 and range between .89 and .98. In general, Phi was parallel to, but slightly higher than the percents.

Which of these statistics a researcher uses appears to depend on the weight he or she gives to the better mathematical properties of Phi as opposed to the greater difficulty in responding to it.

Table 11. Correlations between phi and percent on scored intervals

Behavior	Correlation
Positive	.97
Acknowledge	.98
Negative	.98
General Instruction	.96
Rule	.94
Wrong Responses	.97
Questions	.94
Initiates-to-Teacher	.98

Summary and Conclusions

This study was an attempt to demonstrate that teacher behaviors which have been demonstrated as powerful controllers of student achievement in experimental settings are also influential in the natural environment. Two areas of teacher behavior were considered: behavioral consequences for academic behavior and direct instruction.

The demonstration was most successful in the area of behavioral consequences. Comparisons between groups of teachers indicated that teachers of higher achieving children use more positive social reinforcement and less reprimand than teachers of lower achieving children. These findings suggest that contingent social reinforcement is an

important behavior in improving achievement in the natural environment as well as in experimental situations.

In the area of direct instructions, only one significant difference was found. The rate of General Instruction was lower for Higher Group teachers. It is probable that the relationships with achievement result from Higher Group teachers using more of another effective category rather than less of the general category. What this alternative might be is not indicated by these data. Also noteworthy was the low rate of modeling for all teachers. Despite a fairly large body of literature supporting its effectiveness as a teaching device, it occurred infrequently in the classes in this study. It appears that this is an area where research findings have not had a major impact on teaching practices.

In the area of student behaviors, two differences between groups approached significance. These were Questions and Initiates-to-Teacher. For both, students of the Higher Group teachers made more responses. These results are supported both by past research and by cultural beliefs about optimal learning environments.

A second focus of this study was inter-observer agreement. Three methods of presenting agreement data were compared. Two, Phi, and percentage of scored intervals were found to be highly similar. The third, direct inspection of calibrator data, was viewed as appropriate in this case because of the low rates of many of the categories recorded. The results based on calibrator data were highly similar to those of the primary observer, thus lending confidence to the findings.

APPENDIX A

LETTER TO TEACHERS

To: Teachers and Principals

From: Shirley Cole, M.A.

The study we propose is an exploratory study of teacher style. A substantial amount of research has been carried out on the relation between teacher behavior and student performance. These studies have made use of a variety of standardized tests and rating scales as measures of teacher behavior. In general, these studies have not produced clear interpretable results. One problem with this research is the reliance on indirect measurement. The purpose of the present study is to begin the replication of this research using direct observation of teachers as they normally work in the classroom. Specifically, we hope to delineate the kinds of behaviors teachers engage in and their relative importance in each classroom.

We hope to involve ten teachers in this study. A research assistant would visit each classroom once or twice a week to make systematic observations during the scheduled mathematics period. Assistants would record the kind of interaction between teachers and students as the teacher normally conducts the class. We will try to make a total of twelve visits over the course of the fall semester. On approximately one-third of the visits, a second research assistant will accompany the primary observer in order to determine the accuracy of the data collection. Research assistants will be psychology majors from The University of Arizona and will be trained and supervised by myself.

At the end of each observation we would also like to make a record of the children's success on the most recent math assignment. This would entail providing the observer with the children's papers for about half an hour. They would be returned, of course, before the observer left for the day.

Finally, at the end of the semester we would like to meet with each teacher to obtain a description of the mathematics program.

When all the results have been compiled a central meeting will be held to explain the results. In all data reporting the teachers and children will remain strictly anonymous.

This study will make these demands on teacher time:

First, we will need teachers to fill out a short questionnaire about the time of mathematics period, the preferred days for observations and the basic structure of the class. This will allow us to schedule the observation on the most convenient times for both teachers and observers. This should take about ten minutes.

Second, we will need to meet briefly with each teacher to introduce the observers and make procedural arrangements. This should take no more than fifteen minutes.

Third, observers will need to be able to identify the children in the class by name. In many classes this can be accomplished by providing the observers with a copy of the seating chart. In other classes, name tags may be more appropriate. (If so, these will be provided by the observers.) How this is best done would be worked out at the initial meeting with teachers.

Fourth, observers would make twelve classroom visits. Observers would not interact with the children at all. In general, they will stand or sit quietly in the back of the room writing on their recording sheets. They may need to move occasionally to keep the teacher within ear shot. We find that after the first visit, if teachers ignore the observers children will also. At the end of the mathematics period, the observers would ask the teacher for the most recently completed assignment, preferably the one from that day. He would then go to the teachers' lounge to record the performance on the assignment before returning them to the teacher.

Fifth, we would like to interview each teacher in December to obtain a description of the mathematics program. This should take about half an hour.

Finally, we will hold a central meeting in February or March to present the results to interested teachers and principals.

APPENDIX B

DEFINITIONS OF BEHAVIORS AND DATA COLLECTION PROCEDURE

Teacher Style Study

Definitions of Behaviors

Teacher Behaviors

Describes Materials (DM):

Includes: Any description of a teaching material. The material must be present.

Examples: The problems on page 79 are subtraction.
This is an abacas. You use it to do arithmetic.

Models (M):

Includes: Any time a teacher, herself, does what she wants the children to do.

Examples: Doing problems on the board, saying things for the children to repeat.

Gives Rule (R):

Includes: Any time the teacher tells the child a way of determining whether the response he has made is correct, some kind of test or criterion for accuracy. Also includes any description which includes <u>how</u> to do something.

Examples: You have to borrow if the bottom number is bigger than the top number.
Count over, if it has three places then it's the hundreds column.
You know this is subtraction if it has a minus sign.
If you see the word "of," then use multiplication.
When you need to borrow, take one from the next column and add it here.

Probe (P):

Includes: Any teacher behavior which requires a response from a child, which will allow the teacher to determine whether a child has understood an instruction. These must be questions or commands.

Examples: Do this problem on the board.
If I take 3 away, how many are left?
What do I have to do next?
What does this sign mean?

Excludes: Rhetorical questions such as "Do you really think that is right" or "Would you please hand out the papers." These usually are behavior management techniques, not statements made in response to academic responses.
Also, do not record a teacher only calling on a child.

General Instructions (G):

Includes: Instructions which are not scored as DM or R.

Examples: Today we are going to do the problems on page 73.
Your assignments are due tomorrow.
Bill, pass out the papers.

Positive (+):

Includes: Any teacher response likely to reinforce a response, i.e., verbal praise, tokens, special materials, points, smiley faces.
Teacher indicates response is correct.

Examples: Teacher makes a statement with positive emphasis, "That's right," "Good," "Very nice," "There you go."

Negative (-):

Includes: Any teacher behavior likely to decrease the frequency of the response. Teacher indicates response is wrong.

Examples: Various statements in a harsh tone, "Is that the way we do things here," "No, try again."

Acknowledges Response (Ak):

Includes: Any teacher reaction to a child response which indicates that she knows the child made the response but does not indicate whether it is correct or not. Do not record if accompanied by another code.

Examples: OK. I see.

Repeats Instruction (RI):

Includes: Any instruction by the teacher which repeats the instruction which was given before, without adding new information.

Examples: Remember to borrow in these problems (original instruction).
No, you didn't borrow (score -). Remember to borrow (RI).

Gives Correct Response (GC):

Includes: Any time the teacher tells the child the correct answer, repeats what a child has just said, or answers a child's question.

Examples: Child is counting 1, 2, 3, 5. Teacher says 4.

Child Behaviors

Correct (C):

Includes: Any child response which is adequate and correct from the standpoint of the teacher.

Examples: Child solves math problem correctly.
Child answers question correctly.

Wrong (W):

Includes: Any response which is incorrect or inadequate from the standpoint of the teacher. This includes responses which are only partially correct, if that is unacceptable or if the teacher attempts to correct the child.

Examples: Child counts 1, 2, 3, 5. Child answers a question incorrectly.

No Response (NR):

Includes: Any time a child is called on to respond and says nothing.

Examples: Child does not repeat with the class. Child does not write on his paper when requested.

Questions (Q):

Includes: Any question asked by a child.

Examples: What page are we on?
What does that sign mean?

The following behaviors are recorded on Child Form only.

Initiates to Teacher (I):

Includes: Any time child says something to the teacher or volunteers to respond, whether or not the teacher calls on him/her. Any time child initiates.

Examples: Child goes to teacher's desk. Child waves his hand at teacher.

On-Task (O):

Includes: Any time child is engaged in an appropriate task.

Examples: Child looks at teacher, writes on paper, looks at book, talks appropriately to peers, raises hand, waits for teacher.

Off-Task (F):

Includes: Any time child is passively sitting without working on the assigned task for a prolonged period, any time child actively disturbs others, any time child is out of his seat without permission.

Examples: Child watches another child. Child talks to others inappropriately, child taps pencil on desk loudly, child wanders around room.

No Task Available (Av):

Includes: Any time child has completed the assignment and has not been given another.

Examples: Child is in free play after completing assignment. Class is changing activities.

Excludes: Brief intervals while teacher is waiting for all children to complete a problem.

Teacher Style Study

Data Collection Procedure

Format

Lines One, Two, and Three--the data collection form is divided into five lines. The first and third lines are used to record teacher antecedent and consequent behaviors. The second line is used to record child behaviors.

With a few exceptions the code comprises a mutually exclusive and exhaustive set of categories. All of the relevant behaviors are recorded. In general this will consist of sequences of teacher antecedent, child responses and teacher consequences. Frequently the teacher consequence will also be the antecedent for the next response. The following sequence is an example of how behaviors are recorded:

Teacher: What kind of graph is this? (probe)
Child: A bar graph (correct response)
Teacher: Why do you say that? (probe)
Child: Because of these lines (correct)
Teacher: That's right (positive)

P	
C	C
P	+

Each cell on the form represents 10 seconds of interaction. In most intervals only one sequence of antecedents, response and consequence will occur. In case a second begins, the cell will be divided to indicate this. The sample cell above illustrates this.

At the end of each 10 second interval, recording begins in the next cell. A sequence of behavior may, therefore, be spread over a number of cells. The sequence below is an example of this:

Interval 1
Teacher: What game got the most votes? (probe)
Child: Where are we? (question)
Interval 2
Teacher: Page 27 (gives correct response)
Child: OK, soccer (correct response)
Teacher: Good (positive)

1	2
P	GC
Q	C
	+

A behavior may begin in one cell and continue into the next. In this case, the behavior is recorded in both cells. The following is an example.

Interval 1
Teacher: Today we are going to do subtraction problems
Interval 2
Teacher: They are just like the ones yesterday. What did we do yesterday?
Child: Borrowing
Teacher: Good
Interval 3
Teacher: I'm glad you remembered

1	2	3
G	G P	+
	C	
	+	

Line four--line four is used to record teacher positive and negative responses that are not contingent on the response which was just recorded. Teacher positive and negative may also be recorded on line three, providing that they are in response to the child behavior which has just been recorded. The examples below demonstrate when to record on line three or four.

Teacher: Class you all came in very quietly. That's very good. Get out your books. (positive and general)

G
+

Teacher: John, what does it mean to borrow? (probe)
John: Take one from the 10's column (correct)
Teacher: Good, you remembered (positive)

P
C
+

Line five--this is used to record the names of the children to be observed on child form.

Teacher and Child Forms

(This form can be used to focus on either teacher behaviors or child behavior.)

When recording on the teacher form, observers focus their attention on the teacher and record children's behavior only when they interact with the teacher. Compliance to commands (etc.) directed at the entire group is not recorded.

When recording as child form, observers select one child and focus their attention on that child. Teacher behaviors are recorded only if the teacher interacts with that child or directs his/her behaviors at the whole group. Compliance to commands (etc.) directed to the entire group is recorded. Also, four child behaviors (which are not recorded on teacher form) are added. These are On-Task, Off-Task, No Task Available and Initiates to Teacher. The first three are time-sampled at the break between the intervals, rather than during the interval. The fourth behavior is recorded during the interval.

Children to be observed will be selected randomly from a list of all the children in the class. Each child will be observed for three consecutive intervals. Thus, ten children will be observed in a five minute period.

Data Collection Form

1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6

REFERENCES

Bandura, A. Principles of Behavior Modification. San Francisco: Holt, Rinehart and Winston, Inc., 1969.

Barrish, H. H., Saunder, M. and Wolf, M. M. Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. Journal of Applied Behavior Analysis, 1969, 2, 119-124.

Belgard, M., Rosenshine, B. and Gage, N. L. The teacher's effectiveness in explaining: evidence on its generality and correlation with pupil's ratings and attention scores. In N. L. Gage, M. Belgard, D. Dell, J. Hiller, B. Rosenshine, and W. R. Unruh (Eds.) Explorations of the Teacher's Effectiveness in Explaining. (Technical Report #4) Stanford: Stanford Center for Research and Development in Teaching, Stanford University, 1968.

Brigham, T. A. and Sherman, J. A. An experimental analysis of verbal imitation in preschool children. Journal of Applied Behavior Analysis, 1968, 1, 151-160.

Broden, M., Copeland, G., Beasley, A. and Hall, V. R. Altering student responses through changes in teacher verbal behavior. Journal of Applied Behavior Analysis, 1977, 10, 465-478.

Chadwick, B. A. and Day, R. C. Systematic reinforcement: academic performance of underachieving students. Journal of Applied Behavior Analysis, 1971, 4, 311-319.

Cobb, J. A. Relationship of discrete classroom behaviors to fourth-grade academic achievement. Journal of Educational Psychology, 1972, 63, 74-80.

Cunningham, W. G. The impact of student-teacher pairings on teacher effectiveness. American Educational Research Journal, 1975, 2, 169-189.

Fortune, N. F. A Study of the Generality of Presenting Behaviors in Teaching. U.S. Office of Education, Project #6-8468. Memphis, Tennessee: Memphis State University, 1967.

Gelfand, D. M. and Hartmann, D. P. Child Behavior Analysis and Therapy. New York: Pergamon Press, 1975.

Glynn, E. L. and Thomas, J. D. Effect of cueing on self-control of classroom behavior. Journal of Applied Behavior Analysis, 1974, 7, 299-306.

Hall, R. V., Lund, D. and Jackson, D. Effects of teacher attention on study behavior. Journal of Applied Behavior Analysis, 1968, 1, 1-12.

Hardiman, S. A., Geotz, E. M., Reuter, K. E. and LeBlanc, J. M. Primes, contingent attention and training: effects on a child's motor behavior. Journal of Applied Behavior Analysis, 1975, 8, 399-410.

Harris, W. V. and Sherman, J. A. Use and analysis of the good behavior game to reduce disruptive classroom behavior. Journal of Applied Behavior Analysis, 1973, 6, 405-417.

Hart, B. M. and Risley, T. R. Establishing use of descriptive adjectives in the spontaneous speech of disadvantaged preschool children. Journal of Applied Behavior Analysis, 1968, 1, 109-120.

Hartmann, D. P. Considerations in the choice of inter-observer reliability estimates. Journal of Applied Behavior Analysis, 1977, 10, 103-116.

Hawkins, R. P. and Dotson, V. A. Reliability scores that delude: an Alice in Wonderland trip through the misleading characteristics of inter-observer agreement scores in interval recording. In E. Ramp and G. Semb (Eds.) Behavior Analysis: Areas of Research and Application. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Heath, W. R. and Nielson, M. A. The research basis for performance-based teacher education. Review of Educational Research, 1974, 44, 463-484.

Herman, S. H. and Tramontana, J. Instruction and group versus individual reinforcement in modifying disruptive group behavior. Journal of Applied Behavior Analysis, 1971, 4, 113-119.

Hops, H. and Cobb, J. A. Initial investigation into academic survival skill training direct instruction and first grade achievement. Journal of Educational Psychology, 1974, 66, 548-553.

Hughes, D. C. An experimental investigation of the effects of pupil responding and teacher reacting on pupil achievement. American Educational Research Journal, 1973, 10, 21-37.

Johnson, S. M. and Bolstad, O. D. Methodological issues in naturalistic observation: some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy and E. J. Mash (Eds.) Behavior Change: Methodology, Concepts, and Its Practice. Champaign, Illinois: Research Press, 1973.

Knapczyk, D. R. and Livingston, G. The effects of prompting question-asking upon on-task behavior and reading comprehension. Journal of Applied Behavior Analysis, 1974, 7, 115-121.

Kratochwill, T. R. and Wetzel, R. J. Observer agreement, credibility, and judgement: some considerations in presenting observer agreement data. Journal of Applied Behavior Analysis, 1977, 10, 133-139.

Lahey, B. B. Modification of the frequency of descriptive adjectives in the speech of Head Start children through modeling without reinforcement. Journal of Applied Behavior Analysis, 1971, 4, 19-22.

Long, J. D. and Williams, R. L. The comparative effectiveness of group and individually contingent free time with inner-city junior high school students. Journal of Applied Behavior Analysis, 1973, 6, 465-474.

Medland, M. B. and Stachnik, T. J. Good behavior game: replication and systematic analysis. Journal of Applied Behavior Analysis, 1972, 5, 45-51.

Medley, D. M. and Mitzel, H. E. Some behavior correlates of teacher effectiveness. Journal of Educational Psychology, 1959, 50, 239-246.

Michaels, J. Statistical inference for individual organism research: mixed blessing or curse? Journal of Applied Behavior Analysis, 1974, 7, 647-653.

Packard, R. G. The control of classroom attention: a group contingency for complex behavior. Journal of Applied Behavior Analysis, 1970, 3, 13-28.

Rosenshine, B. and Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed.) Research on Teacher Education. Englewood Cliffs, N.J.: Prentice Hall, 1971.

Rowbury, T. G., Baer, A. M. and Baer, D. M. Interactions between teacher guidance and contingent access to play in developing preacademic skills in deviant preschool children. Journal of Applied Behavior Analysis, 1976, 9, 85-104.

Rowe, W., Murphy, H. B. and DeCsipkes, R. A. The relationship of counselor characteristics and counseling effectiveness. Review of Educational Research, 1975, 45, 231-246.

Sarason, S. B. The Culture of the School and the Problem of Change. Boston: Allyn and Bacon, Inc., 1971.

Schmidt, G. W. and Ulrich, R. E. Effects of group contingent events upon classroom noise. Journal of Applied Behavior Analysis, 1969, 2, 171-179.

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

Soli, S. D. and Devine, V. T. Behavioral correlates of achievement: a look at high and low achievers. Journal of Educational Psychology, 1976, 68, 335-341.

Solomon, D., Bezdek, W. E. and Rosenberg, L. Teaching Styles and Learning. Chicago: Center for the Study of Liberal Education for Adults, 1963.

Suskind, E. C. Questioning and curiosity in the elementary school classroom. Unpublished doctoral dissertation. New Haven, Connecticut: Yale University, 1969.

Walker, H. and Hops, H. Use of normative peer data as a standard for evaluating classroom treatment effects. Journal of Applied Behavior Analysis, 1976, 9, 159-168.

Wallen, N. E. Relationships between Teacher Characteristics and Student Behavior: Part Three. U.S. Office of Education Cooperative Research Project #SAE OE 5-10-181. Salt Lake City: University of Utah, 1966.

White, M. A. Natural rates of teacher approval and disapproval in the classroom. Journal of Applied Behavior Analysis, 1975, 8, 367-372.

Ysseldyke, J. E. Diagnostic-prescriptive teaching: the search for aptitude-treatment interactions. In L. Mann and D. A. Sabatino (Eds.) The First Review of Special Education, Volume 1. Philadelphia: JSE Press, 1973.