HAY, LINDA RUDIN. An Investigation of the Methodological Problems Inherent in the Use of Teachers as Observers of Student Classroom Behavior. (1974) Directed by: Dr. Rosemery O. Nelson. Pp. 88.

The purpose of the present investigation was to systematically evaluate the use of teachers as observers of student classroom behavior. On the basis of prior studies delineating the methodological problems concerning the use of independent observers, it was predicted that teachers would be biased, reactive, and unreliable recorders of student behavior. In addition, it was predicted that the recording of student behavior would effect changes in the teachers' interaction behavior with the students she was observing: observer-mediator reactivity effects.

Independent observers recorded the classroom behavior of four students in each of eight teachers' classrooms for 10 days. In each classroom, two students had been referred by the teacher (referred students) and two other students had been selected by the independent observers (nonreferred students). During observation days 1-5, Interval I, the independent observers recorded the classroom behavior of each student and the teacher's verbal interaction behavior with each student. During observation days 6-10, Interval II, the independent observer continued to observe student and teacher behavior while the teacher concomitantly observed the classroom behavior of one referred and one nonreferred student (experimental condition) in her classroom. The other referred and nonreferred student in each classroom was observed only by the independent observer (control condition) during Interval II.

The results of the study confirmed the predictions of observer reactivity, lack of observer reliability, and observer-mediator reactivity. The teachers were reactive and unreliable observers and teacher

observations of student behavior did effect changes in the teachers' verbal interaction behaviors with these students. Observer bias effects, however, were not evidenced. The implications of these results should be of concern to researchers employing teachers as data collectors in that these methodological confounds may substantially limit the internal and external validity of their experimental findings.

AN INVESTIGATION OF THE METHODOLOGICAL PROBLEMS INHERENT IN THE USE OF TEACHERS AS OBSERVERS OF STUDENT CLASSROOM BEHAVIOR

by

Linda Rudin Hay

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
1974

Approved by

Rosemary O. Nelson

Thesis Adviser

APPROVAL PAGE

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

Thesis Adviser Rosemary O. Nelson

Committee Members Jacquelyn Sachelein

Marilyn T. Erickson

November 25, 1974

Date of Acceptance by Committee

ACKNOWLEDGMENTS

This author would like to express sincere thanks to Dr. Rosemery O. Nelson for her guidance throughout the course of this research investigation. Appreciation is also extended to Dr. Jacquelyn Gaebelein and Dr. Marilyn Erickson for their constructive suggestions.

In addition, appreciation is expressed to Susan Hare and Marsha Haslip for their participation as observers; to the teachers of Rena Bullock and Nathanael Greene Elementary Schools for their cooperation; and to my husband William Hay for his assistance in both data collection and analyses. Finally, this author would like to express special thanks to Dr. William Powers for his exceptional assistance with statistical and computer analyses.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One of the most distinguishing features between behavioral and traditional psychologists is the behaviorists' adherence to a scientific model in both research investigations and in the treatment of the behavior of a single individual (Bandura, 1969; Yates, 1970). In contrast to the more subjective psychological assessment procedures utilized by traditional psychologists to identify "underlying pervasive traits," behavioral psychologists have delineated objective assessment techniques aimed at the direct observation and measurement of the individual's responses to selective environmental stimuli. The behaviorist's recognition of the importance of situational specificity in the assessment of behavior (Mischel, 1973) has contributed to the emphasis on the collection of behavioral data in naturalistic observational settings. These behavioral observation procedures are concerned with the precise delineation of what an individual does in a specific situation and do not attempt to infer either generalized behavioral characteristics or what an individual does in response to other situational stimuli.

The majority of behavioral research investigations have relied upon independent observers to record the behavior of specific individuals. Some of the behaviors that independent observers have recorded include the behavior of children in the school classroom (Becker, Madsen, Arnold, & Thomas, 1967; Bersoff & Ericson, 1972; Cobb & Ray, 1971), family interaction behaviors in the home (Patterson, Ray, &

Shaw, 1969), and the behavior of individuals in psychiatric hospital environments (Ayllon & Azrin, 1964). Typically, due to the complexity of behavior, only the occurrence of particular responses is recorded. Each response category is operationally defined such that the observer can determine whether or not the response has occurred. The individual's behavior is recorded both prior to and during the implementation of specific treatment procedures. This initial identification of a baseline rate of behavior preceding the initiation of treatment procedures enables the behaviorist to continually monitor the treatment effectiveness.

Although these observational techniques have increased the objectivity of psychological assessments, researchers have grown increasingly aware of the methodological problems inherent in the use of these behavioral procedures (Johnson & Bolstad, 1973; Lipinski & Nelson, 1974; O'Leary & Kent, 1973; Patterson & Harris, 1968). Three eminent areas of concern are observer bias, observer reactivity, and observer reliability. Lipinski and Nelson (1974) define observer bias as consistent changes in the observer's recording behavior in response to factors other than the observee's behavior. Observer reactivity refers to actual changes in the observee's behavior in response to being observed. Observer reliability includes methodological problems resulting in a lack of consistency in the measurement of behavior.

Factors that contribute to observer bias include the awareness that reliability is being assessed and the knowledge of predicted results. Research investigations have dramatically demonstrated the

importance of keeping observers unaware of the specific assessment of observer reliability. Reid (1970) found that observer reliability dropped from a median of .75 to .51 when observers were led to believe that reliability was no longer being assessed. Reid and DeMasters (1972) found that the level of reliability fell from .82 on the last day of observer training to .27 on the first day of covert reliability assessment. Romanczyk, Kent, Diament, and O'Leary (1973) reported that not only did reliability drop when observers were unaware that reliability was being assessed, but also that knowledge of which observer was assessing reliability significantly increased the reliability score. They suggest that pairs of reliable observers may adjust their definitional criteria for scoring or "drift" away from the original codes over time in order to maintain high reliability scores.

These studies clearly exemplify the importance of covert methods of reliability assessment. Johnson and Bolstad (1973) suggest the use of a random-check reliability procedure in research investigations when covert assessment of reliability is impossible or impractical. Taplin and Reid (1973) compared observer reliability scores obtained when observers were led to believe that reliability would not be assessed (no-check), reliability would be assessed only at predictable times (spot-check), and reliability would be assessed at random and unpredictable times (random-check) throughout the experiment. All three groups of observers showed a significant decrease in the level of reliability immediately following the shift from training to actual data collection. The random-check technique, however, maintained a nonsignificant but

consistently higher level of reliability across data collection sessions than either the spot-check or no-check reliability assessment procedures. In addition, reliability scores in the spot-check condition were significantly higher on days when the observers were aware that reliability was being assessed than on the days when they were led to believe that reliability was not being assessed.

Research investigations that have intentionally manipulated observer knowledge of predicted results have reported contradictory findings. Several studies report significant changes in the observer's data as a result of induced observer expectations of treatment effects (Kass & O'Leary, 1970; Rosenthal & Fode, 1963; Scott, Burton, & Yarrow, 1967). More recent studies, however, have failed to confirm these results (Kent, O'Leary, Diament, & Dietz, 1974; Skindrud, 1971). Kent et al. (1974) suggest that the results of the Kass and O'Leary (1970) study may have been confounded by observer "drift" from the behavioral code definitions. Although Kent et al. (1974) did not find significant observation biases in the recorded data, the observer's verbal reports of the observee's behavior were concordant with the induced experimental expectations. O'Leary, Kent, and Kanowitz (1974) did find that it was possible to shape data concordant with experimenter expectations by providing experimenter feedback to the observers indicating how well the data compared with the experimenter's hypotheses.

In addition, Mash and Makohoniuk (1974) have suggested that observers may form their own expectations as to the behavior of the individuals that they are observing. Mash and Makohoniuk (1974) intentionally

manipulated observer expectations as to the predictability of behavioral sequences to be observed. The accuracy of observers who were led to believe that the behavioral sequences contained predictable behavioral patterns was significantly lower than the accuracy of observers not informed as to the predictability of the behaviors being observed or specifically instructed that there were no predictable patterns.

In spite of the controversial nature of these experimental findings, Johnson and Bolstad (1973) strongly recommend keeping the observer's knowledge of both the predicted results and experimental treatment phases (baseline-treatment-follow up) at a minimum in order to minimize the potential of observer bias effects. The presence of observer bias effects may markedly reduce the internal validity of both between-subject and within-subject experimental designs. In between-subject designs, when different pairs of observers are assigned to observe individuals in different treatment conditions, the problems of observer "drift" may contribute unsystematic and random error to the experiment thereby reducing the chance of detecting important phenomena. Failures to reject statistical hypotheses as a result of error variance may likely be attributed to theoretical rather than methodological concerns. Even more importantly, knowledge of predicted results and/or treatment conditions may result in a systematic alteration of the recorded data of the behavior of different treatment groups leading to the confounding of observer bias effects and the effects of the independent variables. The results of within-subject or time series experiments may be confounded if observers "drift" in their application

of a behavioral code over time resulting in incomparable data from different experimental treatment phases or conditions. Likewise, knowledge of the experimental treatment phase (baseline-treatment-follow up) and/or predicted results may systematically confound experimental findings (O'Leary, 1973).

In addition to the concern with observer bias effects, researchers have also investigated the effects of observer reactivity on the behavior of the individual being observed. In order to obtain an objective assessment of an individual's behavior, observations must be made in such a way as to leave the natural environment unaltered. Researchers have employed hidden mechanical devices and one-way mirrors in order to reduce the effects of being observed on the behavior of the observed individual (Webb, Campbell, Schwartz, & Sechrest, 1966). Frequently, however, these procedures are impractical and the observer must be visible as he records the individual's behavior. Typically, under these conditions, the observers are instructed to "fade into the walls" (Becker, Madsen, Arnold, & Thomas, 1967) and to extinguish interactions with the individuals being observed (O'Leary, Romanczyk, Kass, Dietz, & Santogrossi, 1971) in order to become as neutral an environmental stimulus as possible. It has generally been recommended that the observers be present in the observational setting long enough for the individuals being observed to "habituate" to their presence prior to the initiation of actual data recording sessions (Patterson & Harris, 1968).

Systematic research investigations, however, have indicated that these procedures may be insufficient in reducing observer reactivity

effects. Although several studies have failed to demonstrate that the presence of an independent observer effects changes in the behavior of the individuals being observed (Bales, 1950; Callahan & Alevizos, 1974), the presence of an independent observer has been shown to significantly alter the behavior of nursery school children (Arsenian, 1943), delinquent children (Polansky, Freeman, Horowitz, Irwin, Papanis, Rappaport, & Whaley, 1949), visitors in an art museum (Bechtel, 1967) and family interaction behaviors (Patterson & Reid, 1969) in the natural environment. In addition, Grimm, Parsons, and Bijou (1972) found that school children did not habituate to the presence of observers but continued to look at them with a high frequency even after the observers had been present every day in the classroom for a period of six months. Candland, Dresdale, Leiphart, and Johnson (1972) found that the presence of human observers effected changes in the frequency of certain behaviors exhibited by nonhuman primates following as much as three years of contact with human observers.

Similarly, Johnson and Bolstad (1973) have suggested that individuals who are aware that they are being observed may "fake" their behaviors to conform with the behavior considered appropriate in a given situation. Roberts and Renzaglia (1965) found that clients made more favorable self-references when they knew that their therapy sessions were being recorded than when they were not aware that the sessions were recorded. Patterson and Reid (1969) reported that the presence of an independent observer effected an increase in the frequency of positive social reinforcers among family members and a decrease in

negative reinforcers. Johnson and Lobitz (1972) demonstrated that the demand characteristics of being observed could alter the behavior of the observed individuals. In this study, parents were able to systematically vary the frequency of deviant behaviors exhibited by their children in the desired direction by modifying their own behavior during observation sessions.

Both the internal and external validity of between-subject and within-subject experiments may be affected by observer reactivity. In between-subject experiments, unequal observer reactivity effects across treatment conditions or an interaction effect between observer reactivity and the dependent variable may reduce the internal validity and confound the experimental results. On the other hand, observer reactivity may be equal across treatment conditions but limit the external validity or generalizability of experimental findings (Johnson & Bolstad, 1973; Webb et al., 1966). The internal validity of within-subject and time series experiments may be affected by changes in observer reactivity if individuals adapt over time (Patterson & Harris, 1968). However, even if the behavior of the observed individuals in response to being observed remains constant over time, the external validity or generalizability of experimental results may nevertheless be impaired.

Finally, a lack of observer reliability in the recording of behavior by an independent observer may contribute to the distortion of experimental findings. Reliability of behavioral observations refers to the extent of agreement between the data recorded by two independent observers using the same recording procedure while viewing the same

behavioral sequence. Johnson and Bolstad (1973) suggest that the behaviorist's lack of concern with the reliability or consistency of measurement may reflect misinterpretations of Mischel's (1968) exposition on the inconsistency of behavior over time and across environmental situations. The inconsistency of human behavior, however, does not relieve the researcher from the requirements of reliability of measurement. Although the behavior of the same individual may vary as a result of minor alterations in environmental stimuli or with the passage of time, the recording of an individual's behavior in a delineated situation at a specific time must yield reliable or replicable scores. Instrument decay and the "faking" of observational data may reduce the reliability of behavioral scores.

Instrument decay (Campbell & Stanley, 1966) refers to the tendency for measurement accuracy to deteriorate over time. In the majority of behavioral investigations, observers are trained to an initial high level of inter-observer reliability and are subsequently assigned to various environmental settings where their observational behavior is no longer supervised. Under these conditions, the accuracy of human observers may decay as a result of fatigue, boredom or forgetting. In addition, observers may "fake" the data by scoring behavioral observation code sheets without actually observing the observee's behavior. All of these factors may function to reduce the reliability of the data and thus the validity of experimental results.

Several researchers have proposed using mediators as observers in order to circumvent the methodological problems engendered by the use

of independent observers (Foster, Keilitz, & Thomas, 1974; Kubany & Sloggett, 1973; Surratt, Ulrich, & Hawkins, 1969). A mediator is an individual who controls some of the target individual's reinforcers and who will be responsible for implementing treatment procedures. Researchers have suggested that the use of mediators as observers might reduce those problems attributed to the addition of an independent observer to the environment. Furthermore, the use of mediators as observers may be more economical and convenient since these individuals are already a part of the observee's environment. Many research investigations have utilized mediators as data collectors prior to treatment implementation (Hall, Christler, Cranston, & Tucker, 1970; Hall, Fox, Willard, Goldsmith, Emerson, Owen, Davis, & Porcia, 1971; Kubany & Sloggett, 1973; McAllister, Stachowiak, Baer, & Conderman, 1969). It is conceivable that the same methodological problems inherent in the use of independent observers, may function to jeopardize the objectivity of behavioral assessments when mediators are employed as data collectors. Specifically, the same observer bias factors, awareness that reliability is being assessed and the knowledge of predicted results, that contribute to changes in the data recorded by independent observers may affect the data recorded by mediators. When mediators are employed as data collectors covert reliability assessment is frequently impractical since the majority of naturalistic observational settings are not equipped with one-way mirrors and thus a second observer must be present. Although the experimenter may be able to conceal knowledge of the predicted results, mediators may hold personal expectations as to the

behavior of the observee prior to and during the experimental manipulations (Mash & Makohoniuk, 1974). Furthermore, inasmuch as the mediator is typically the treatment implementer, it may be impossible to withhold knowledge of the experimental treatment phases.

Although numerous research investigations have demonstrated the reactive effects of being observed by an independent observer (Arsenian, 1943; Bechtel, 1967; Patterson & Harris, 1968; Polansky et al., 1949), the reactive effects of observations when mediators are employed as data collectors have not been systematically investigated. Inasmuch as the goal of behavioral assessment is to objectively specify what an individual does in response to specific environmental stimuli, researchers have been aware that the addition of an independent observer may substantially limit the generalizability of their experimental results. Several researchers have suggested the use of mediators as observers in order to eliminate the reactive effects of being observed on the behavior of the observed individual (Foster et al., 1974; Kubany & Sloggett, 1973). Unlike the independent observer, the mediator is already present in the individual's environment and thus it has generally been assumed that problems of observer reactivity will be minimal (Patterson & Harris, 1968; Patterson & Reid, 1969). When the mediator records the behavior of an individual, however, he is engaging in behaviors that are variant to his usual functions in the environment. Thus, the use of mediators as data collectors may still alter the environment in such a way as to systematically modify the behaviors of the individuals being observed.

A study reported by Surratt, Ulrich, and Hawkins (1969) demonstrates the possible reactive effects of observations by mediators on the behavior of the observed individuals. This investigation employed a fifth-grade public school student, as both treatment implementor and data recorder, to modify the study behavior of four first-grade students. The results indicated that although study behavior increased when the observer was present, the increases were only partially maintained when the observer was not present. Surratt et al. (1969) propose that the fifth-grade student became a discriminative stimulus for appropriate study behavior. In addition, reports of "baseline cures" or decreases in the frequency of target behaviors that occur during baseline observation, prior to the initiation of the treatment procedures (Crowder & Willis, 1972) suggest that observations made by mediators may be reactive.

Furthermore, the use of mediators as observers does not eliminate the problem of observer unreliability. The mediator may be as prone to instrument decay and as likely to "fake" observational data as the independent observer. Whereas the independent observer may "fake" observational data in order to receive his paycheck, course credit, or experimenter approval, the mediator may fear the withdrawal of assistance or experimenter disapproval. In addition, the mediator may be less accurate than the independent observer in the scoring of behavioral observations for unlike the independent observer, whose primary function is to record the behavior of designated individuals, the mediator must assume additional roles in the environment. Independent

observers typically receive extensive observational training in the use of behavioral codes, until a demonstratably high level of observer reliability is achieved. Such extensive training is practically never possible when mediators are employed as data collectors.

Finally, the use of mediators as opposed to independent observers may engender additional methodological problems that may function to reduce both the internal and external validity of experimental results. It is conceivable that merely observing the behavior of an individual may effect changes in the observer's behavior. Recording the behavior of an individual necessitates that the observer attend to that individual's behavior. If this attention is contingent upon the performance of a specific behavior, it may serve to systematically alter the frequency of that behavior. In addition, by attending to an individual's behavior, the mediator may recognize antecedent conditions that when varied may effect changes in the frequency of certain behaviors. Changes in the observer's behavior with respect to his function as a mediator in response to the recording of the observee's behavior will subsequently be referred to as observer-mediator reactivity.

Forehand (1973) described a case study in which the behavior of spitting decreased significantly following three days of baseline observation by the teacher. Forehand attributed the decrease in the target behavior to changes in the teacher's behavior. Crowder and Willis (1972) also hypothesized that "baseline cures" were a result of changes in the teachers' responses to the target behaviors. These studies, although reporting desired changes in the frequency of undesirable

behaviors, suggest that the behavioral observations recorded by mediators may be confounded by observer-mediator reactivity effects.

The use of mediators as observers seems to parallel the use of independent observers in that both observational procedures entail the systematic observation and recording of one individual's behavior by another. In view of the experimentally demonstrated methodological problems inherent in the use of independent observers, it seemed probable that the use of mediators as data collectors would engender procedural problems.

In the present investigation, eight teachers observed the behavior of two students (experimental condition) in each of their respective classrooms. Prior to and during teacher observations, independent observers also recorded the classroom behavior of these students in the experimental condition. In addition, the independent observers recorded the behavior of two other students (control condition) in each teacher's classroom in order to control for any extraneous factors apart from teacher observations that might have accounted for changes in student behaviors. The independent observers also concomitantly recorded the teacher's interaction behaviors with each of the experimental and control condition students. One experimental and one control condition student in each class were referred by the teacher.

Specifically, the following four hypotheses were proposed:

1. Teacher observations would be biased. The data recorded by the teacher would indicate a significant difference in the percent of appropriate behavior for the referred and nonreferred students. The

behavior of the referred students would be recorded as less appropriate than the nonreferred students since the teachers personally selected the referred students on the basis of a high frequency of "off-task" behavior.

2. Teacher observations would be reactive as evidenced by a significantly greater change in the behavior of the students observed by the teacher (experimental condition) than for students not observed by the teacher (control condition).

3. The data recorded by the teacher would be an unreliable measure of student classroom behavior. Reliability (inter-observer agreement) between the data recorded by the teacher and independent observer would be less than 85% agreement.

4. Observations of student behavior recorded by the teacher would effect changes in the teacher's verbal interaction behaviors with these students: observer-mediator reactivity effects. A significantly greater change in teacher verbal interaction behavior with the students observed by the teacher (experimental condition) than with the students not observed by the teacher (control condition) was predicted.

CHAPTER II

METHOD

Subjects

Teachers. Two first-grade, two second-grade, and four third-grade teachers participated in this experiment. Teachers were requested to take part in the project by the author (School Behavior Management Specialist) under the guise that the purpose of the study was to delineate effective treatment techniques for dealing with students who frequently exhibited "off-task" classroom behavior. The teachers were not informed as to the actual experimental hypotheses until the study was completed. All teachers asked agreed to participate. Prior to the initiation of the project, each teacher signed a consent form confirming her agreement to participate in the project to completion.

Students. Four male elementary school students (grades 1-3) from each of the eight teacher's classrooms participated in this experiment. Of these four students, two students were referred by the teacher and two students (nonreferred) were selected on the basis of observations by independent observers. One referred student and one nonreferred student in each classroom were subsequently randomly assigned to the experimental and control conditions. Experimental subjects were students whose classroom behavior was systematically observed by the teacher. Control subjects were not observed by the teacher and provided a control for any concurrent changes in the classroom situation that may have effected changes in student classroom behaviors during the experiment.

Referred students were selected on the basis of the teacher's subjective evaluation of their classroom behavior. Each of the eight teachers was requested to submit the names of two male students who were frequently "off-task" during independent study classroom activities. "Off-task" behavior was defined as behavior other than task-related, such as daydreaming, talking without permission, doodling on paper, and out of seat.

Nonreferred students were selected on the basis of observations of classroom behavior by independent observers. An independent observer informally observed the remaining nonreferred male students for a few minutes in each teacher's classroom during an activity requiring each student to work independently at his desk. On the basis of these observations, the five male students who appeared to exhibit the greatest amount of "off task" behavior were selected. Subsequently, the classroom behavior of each of these students was observed and recorded by an independent observer for 10 minutes. The independent observer utilized a time-sampling observation procedure (5-second intervals) to record the classroom behavior of each student. During each 5-second interval, the student's behavior was scored as either on-task (+) or off-task (-) by circling the appropriate code on a pre-coded observation sheet (Appendix A). The student's behavior was recorded as on-task (+) if the student was attending to the task assigned by the teacher for the entire 5-second interval. The student's behavior was recorded as off-task (-) if the student engaged in any behavior other than the assigned task for any part of the interval. If the observer was uncertain

whether or not the student was on-task (+), the interval was scored as off-task (-). Thus, a total of 120 observations of classroom behavior were recorded for each student. On the basis of these observations, the two students in each classroom demonstrating the greatest percent of "off-task" behavior during this time were selected. A second independent observer simultaneously recorded the behavior of these students during 25 percent of the observation time in order to establish the reliability of the observational procedures. Observer reliability was 98% as determined by the number of agreements divided by the number of agreements and disagreements.

In summary, a total of eight teachers and 32 male elementary school students participated in this study. Four groups of eight subjects were delineated: experimental referred; experimental nonreferred; control referred; and control nonreferred.

Behavior Definitions and Recording by Independent Observers

Observer Training. Two female undergraduates and one male and one female (author) graduate psychology students served as independent observers. The undergraduate observers received course credit for their participation.

The observers were trained individually by the author. Each observer was given written instructions describing the observation procedure and behavior codes in detail. Each observer and the author randomly selected students and recorded their behavior for 15 minutes until an inter-observer reliability (agreement) score of at least 90% agreement

was obtained on two consecutive observations. The author and observer discussed the procedure and behavior codes following each observation. All observers obtained the 90% inter-observer reliability criterion within four observations (one hour).

Recording Procedure. The independent observers used a modified version of a time-sampling technique suggested by Hamerlynck and described by Kubany, Block, and Sloggett (1971) to record student and teacher behaviors. Each student was observed for 15 minutes each day for 10 consecutive school days. During the 15-minute observation session, the observer glanced over at the student every 15 seconds and scored his behavior at that instant as either appropriate (A), passive (P), or disruptive (D) as defined in Appendix B. In total, 60 observations of the student's classroom behavior were recorded during each 15-minute observation session. The approximate percent of time that the student was appropriate, passive, and disruptive was determined by dividing the total number of A, P, and D recorded by the total number of intervals (60) and multiplying by 100.

During the remainder of each 15-second interval, verbal behavior exhibited by the teacher directly toward the observed student was recorded. Teacher verbalizations addressed to the entire class were not recorded. Only the following types of verbal behaviors, as defined in Appendix C, were recorded: praises (+), prompts (0), and criticism (-). If the observer was unable to hear the words spoken by the teacher, the verbalization was recorded as a prompt. Each of these categories was

scored only one time during each 15-second interval regardless of the number of verbal behaviors observed. Furthermore, the behavior was recorded only in the interval in which it began. A sample code sheet is shown in Appendix D.

A "random check reliability technique" suggested by Johnson and Bolstad (1973) was used to determine the reliability (inter-observer agreement) of observations recorded by the independent observers. This procedure required that two observers be recording in the same classroom during all recording sessions. The author was present as an observer during all sessions throughout the experiment. All reliability scores were determined between the data recorded by the author and one other independent observer. The observation recording schedules of the two independent observers were pre-arranged such that the two observers were recording the behavior of the same student 50% of the time at random times during each observation period. Using this method, the observers were unaware of the specific times that reliability was being assessed. Johnson and Bolstad (1973) report that this technique not only reduces the observer bias problem resulting from the knowledge that reliability is being assessed, but may also increase the accuracy levels and stability in the observation recording session in general. In addition, all observers, except for the author, were kept blind as to the referral status of each student and whether the student was in the experimental or control condition.

Procedure

Each teacher was asked to select one hour of the day for participation in the research project and during which time independent observers could be present in the classroom. The students were to be assigned independent study activities during this hour. The observation time was held constant each day.

During the hour selected by the teacher, independent observers recorded the classroom behavior of each of the four students (two in the experimental condition, two in the control condition) for 15 minutes a day for 10 consecutive school days. Teacher verbalizations addressed specifically to these students were also recorded during this time. The 10 days of observations were divided into two intervals. Interval I, days 1-5, provided a baseline rate of each student's behavior and the respective teacher's verbalizations to that student prior to the initiation of the teacher's observations of the students in the experimental condition. During Interval II, days 6-10, the teacher was instructed to observe and record the behavior of the two students in the experimental condition: one referred and one nonreferred student. The teacher was told that observation of the nonreferred student would enable a comparison to be made between the behavior of the referred student and a typical student in her class. The teacher was led to believe that the nonreferred student had been selected at random. Thus during Interval II, both the teacher and the independent observers concomitantly recorded the classroom behavior of one referred and one nonreferred student during some portion of the observation hour each day. In addition,

the independent observer alone continued to record the classroom behavior and teacher verbalizations addressed to the students in the control condition: one referred and one nonreferred student.

Teacher Observation Procedure

Each teacher was given written instructions describing the observation technique in detail (Appendix E). The author discussed the observation procedure with each teacher and emphasized the importance of recording objective data. The teacher was cautioned against modifying her behavior in any way during this time.

The observation procedure that was utilized by the teacher to record the classroom behavior of the two students in the experimental condition has been outlined in detail by Kubany and Sloggett (1973). Specifically, the observation sheet (Appendix F) specified a starting time and the number of minutes for which a kitchen timer would be set. The timer was set by an independent observer. When the timer ran out and the bell rang, the teacher was to glance over and record the behavior of each of the students at that instant as appropriate (A), passive (P), or disruptive (D). The timer was immediately reset by the independent observer for the next specified interval of time. The length of the intervals between observations was pre-determined on a variable interval (VI) 4-minute schedule. In total, the teacher recorded the behavior of the two students 15 times during the observation hour each day. The percent appropriate, passive, and disruptive behavior exhibited by each of the students was calculated daily be dividing the total number

of A, P, and D recorded by the total number of intervals (15) and multiplying by 100.

Prior to the initiation of observations, the teacher informed the class that a bell would ring occasionally during the day. The teacher explained that the bell was a memory device to remind her of things that she wanted to do. In addition, the teacher was instructed to ignore student references to the timer.

CHAPTER III

RESULTS

Reliability (Observer Agreement) for Independent Observers

In order to determine the reliability of observations recorded by the independent observers, inter-observer agreement was determined for 38% of the 15-minute observation sessions of student and teacher behaviors. Two methods of determining inter-observer reliability were calculated. A Pearson product-moment correlation coefficient (Method I) was calculated for each student and teacher behavior, on the scores obtained during each 15-minute observation session during which two observers recorded the behavior of the same student. The mean inter-observer reliability score over all student and teacher behaviors, obtained using Method 1, was 97% agreement. Johnson and Bolstad (1973), however, point out that a correlation reflects only the level of agreement on the total obtained and does not necessitate any agreement on specific events. Therefore, Method II, exact agreement, was also calculated by dividing intervals of agreement by intervals of agreement plus disagreement. Repp, Deitz, Boles, Deitz, and Repp (1974) reported that the exact agreement method provides a more conservative measure of inter-observer reliability. The mean inter-observer reliability score over all student and teacher behaviors, calculated using Method II, was 92% agreement, confirming the findings of Repp *et al*. The mean inter-observer reliability for each student and teacher behavior, as determined by each method, is presented in Table 1.

The author was present as an observer during all sessions in which inter-observer reliability was assessed. The author's observation data were therefore used in all subsequent analyses.

Observer Bias

It was hypothesized that teacher observations of student classroom behavior would be biased: the teacher would record the behavior of the referred student as less appropriate than the behavior of the nonreferred student. This was predicted because the teacher had selected the referred student on the basis of a subjective evaluation of a high frequency of off-task behavior whereas the teacher was informed that the nonreferred student had been selected at random and represented a "typical" student in the class.

Prior to the determination of observer bias effects, an analysis of variance was calculated using the independent observer's data in order to demonstrate that the referred and nonreferred students did not differ significantly in the percent of appropriate behavior exhibited during Interval II. The results of the analysis, included in the description of the results for observer reactivity, confirmed this preliminary assumption. The main effect for referral status, the condition x referral status interaction and the condition x referral status x interval interaction were not significant. Thus, the independent observer's data showed that the referred and nonreferred students in the experimental condition did not differ significantly in the percent of appropriate

behavior exhibited during Interval II, when the teachers were recording their classroom behavior.

In order to determine if the teacher recorded a significant difference between the referred and nonreferred students in the percent of appropriate behavior, a 2x5 repeated measures analysis of variance was calculated using the data recorded by the teachers during Interval II. An arc sin transformation of percentage data was performed, prior to the analysis, in order to meet the assumptions of the analysis of variance. The analysis is summarized in Table 2. The results indicated no significant main effect for referral status ($F = 0.1545$; $df = 1, 7$), thus the percent of appropriate behavior for the referred and nonreferred students as recorded by the teachers did not differ significantly. The results failed to support the hypothesis of observer bias effects. In addition, no other main effects and no interactions were significant.

Observer Reactivity

It was hypothesized that teacher observations would be reactive, effecting changes in the classroom behavior of the students in the experimental condition whose behavior was observed by the teacher during Interval II. In order to test this hypothesis, three 2x2x2x5 repeated measures analyses of variance were calculated for the percent of appropriate, passive, and disruptive behavior exhibited during Interval I and Interval II for students in each of the four conditions. A significant condition x interval interaction was predicted: the classroom

behavior of the students in the experimental and control conditions would not differ significantly in Interval I, but would differ significantly in Interval II, when the teachers were recording the behavior of the students in the experimental condition.

The independent observer's data were used in all analyses. An arc sin transformation of percentage data was carried out prior to the analyses in order to meet the assumptions of the analysis of variance.

Appropriate Behavior (see Table 3). The results of the analysis revealed no significant main effects at .05 level. There was a significant condition x referral status interaction ($F = 4.57$; $df = 1, 28$; $p < .05$). A Newman-Keuls test of the means of the interaction indicated that the nonreferred students in the control condition exhibited significantly less appropriate behavior than the referred students in the control condition during Interval I and Interval II at the .05 level. The referred and nonreferred students in the experimental condition did not differ significantly in the percent of appropriate behavior exhibited during Interval I and Interval II. In addition, the referred students in the experimental and control conditions and the nonreferred students in the experimental and control conditions did not differ significantly in appropriate behavior.

It had been predicted that the students in the experimental condition would differ significantly from the students in the control condition in the percent of appropriate behavior exhibited during Interval II, when the teachers were observing their classroom behavior. The

condition x interval interaction was not significant, however, suggesting that observations by the teachers were not reactive for appropriate classroom behavior.

Passive Behavior (see Table 4). The analysis indicated a significant main effect for referral status ($F = 8.36$, $df = 1.28$; $p < .01$). A main effect for referral status had not been predicted. The results suggested, however, that the nonreferred students selected by the independent observers as off-task were significantly more passive throughout Interval I and Interval II than the referred students selected by the teachers on their subjective evaluation of a high frequency of off-task behavior.

No other main effects or interactions were significant. The condition x interval interaction was again not significant: the students in the experimental and control conditions did not differ significantly in the percent of passive behavior exhibited in Interval II. Thus the analysis failed to support the prediction of observer reactivity effects for passive behavior.

Disruptive Behavior (see Table 5). The analysis revealed no significant main effects or interactions at .05 level. It had been predicted that teacher observations would be reactive, effecting changes in the percent of disruptive behavior exhibited by the experimental students in Interval II when the teachers were recording these students' behavior. The results failed to confirm the hypothesis since the condition x interval interaction was not significant.

In summary, the results of these analyses revealed no significant differences between students in the experimental and control conditions in the percent of appropriate, passive, or disruptive behavior exhibited during Interval I or Interval II. Thus, the analyses did not support the prediction of observer reactivity effects. The analyses did suggest that the nonreferred students, selected as off-task by the independent observers, were more passive than the referred students, selected as off-task by the teachers, as evidenced by a significant effect of referral status for passive behavior. In addition, the significant condition x referral status interaction for appropriate behavior indicated that the nonreferred students who were randomly assigned to the control condition exhibited less appropriate behavior than the referred students randomly assigned to the control condition throughout Interval I and Interval II.

A closer scrutiny of the data, however, suggested that the lack of observer reactivity effects (condition x interval interaction) might have been the result of the direction of behavior change exhibited by each student from Interval I to Interval II. Over all 32 students, 16 showed an increase and 16 a decrease in appropriate behavior; 15 students exhibited an increase and 17 students a decrease in passive behavior; 15 students showed an increase and 17 students a decrease in disruptive behavior from Interval I to Interval II. The magnitude and direction of change in appropriate, passive, and disruptive behavior from Interval I to Interval II is presented in Table 6. Three condition x referral status analyses of variance for the absolute change from Interval I to

Interval II in appropriate, passive, and disruptive behavior respectively were calculated. A significant condition main effect was predicted: teacher observations would be reactive, effecting a significantly greater change, increase or decrease, from Interval I to Interval II in the classroom behavior of students in the experimental condition, observed by the teacher during Interval II, than students in the control condition whose behavior was not observed by the teacher in Interval II.

Appropriate Behavior (see Table 7). The results of the analysis revealed a significant condition main effect ($F = 4.44$; $df = 1, 28$; $p < .05$) indicating that the students in the experimental condition exhibited a significantly greater change in appropriate behavior from Interval I to Interval II than students in the control condition. The results confirmed the prediction that teacher observations would be reactive, effecting changes in appropriate behavior. No other main effects or interactions were significant.

Passive Behavior (see Table 8). The analysis indicated a significant condition x referral status interaction ($F = 5.66$; $df = 1, 28$; $p < .05$). There were no other significant main effects or interactions.

The condition x referral status interaction is depicted graphically in Figure 1. The means of the interaction were compared via a Newman-Keuls test. The results indicated that the nonreferred students in the experimental condition exhibited a significantly greater change in passive behavior than the nonreferred students in the control condition. No other mean comparisons were significant. The results suggest

that for the nonreferred students, selected on the basis of observations by the independent observer, teacher observations were reactive, effecting changes in passive behavior. Teacher observations of the referred students, selected by the teacher, were not reactive for passive behavior.

Disruptive Behavior (see Table 9). The condition x referral status interaction was significant at .10 level ($F = 3.09$; $df = 1, 28$; $p < .10$). No other main effects or interactions were significant.

The condition x referral status interaction is represented graphically in Figure 2. A Newman-Keuls means comparison test indicated that the referred students in the experimental condition whose behavior was observed by the teacher in Interval II showed a significantly greater change in disruptive behavior than the referred students in the control condition at .05 level. No other mean comparisons were significant. The results indicate that for the referred students, selected by the teacher, teacher observations were reactive for disruptive behavior. The referred students observed by the teacher during Interval II changed more than the referred students not observed by the teacher during Interval II in the percent of disruptive behavior exhibited from Interval I, during which only the independent observers were recording their behavior, to Interval II, during which both the independent observers and teachers recorded their classroom behavior.

Observer Reliability (Teacher-Independent Observer Agreement)

In order to determine if the teacher's recorded observations were reliable, teacher-independent observer agreement was calculated for all observation sessions during which the teacher and independent observer concurrently recorded the classroom behavior of students in the experimental condition. The independent observer's data was considered to be a reliable measure of student behavior because of the high inter-observer reliability scores obtained between the independent observers for all three student behaviors recorded. Therefore, low inter-observer agreement (below 85% agreement) between the data recorded by the teacher and independent observer was considered indicative that the teachers were unreliable observers. Two methods of computing inter-observer agreement were calculated.

Exact Agreement. The concurrent recording of the classroom behavior of students in the experimental condition by the teacher and independent observer resulted in an average of four simultaneously recorded observations during each 15-minute observation session. Inter-observer reliability was calculated for each category of student behavior by dividing the number of agreements by the total number of agreements plus disagreements. The results are presented in Table 10. Reliability scores for all three student behaviors were well below 85% agreement confirming the hypothesis that teacher observations would yield unreliable measures of student classroom behavior.

Pearson Product-Moment Correlation Coefficient. A Pearson product-moment correlation coefficient was calculated for the percent of appropriate, passive, and disruptive behavior recorded by the independent observer and the teacher during each 15-minute observation session in which both the teacher and independent observer simultaneously recorded the classroom behavior of the same student. The percent of time the student engaged in each category of student behavior was determined by dividing the number of recorded observations of each category, respectively, by the total number of observations recorded during each session. The total number of observations recorded by the independent observer during each 15-minute observation session was 60 (fixed-interval 15-second observation schedule). The total number of observations recorded by the teacher was three or four per observation session (variable interval 4-minute observation schedule). Kubany, Bloch, and Sloggett(1971) recalculated data recorded on a fixed-interval 15-second schedule by independent observers of student disruptive behavior as if it had been recorded on a variable-interval 4-minute schedule by computing the percentage of disruptive behavior from the four or five juncture points. They reported that the percents of disruptive behavior as calculated by each method were highly similar. Furthermore, elementary probability theory predicts that randomly-selected observations should yield reasonably reliable observational data (Hays, 1963). Thus, low inter-observer reliability scores would most likely not be a result of the different number of observations recorded by the teachers and independent observers during the 15-minute observation session.

The results of the Pearson product-moment correlation are presented in Table 10. The inter-observer reliability scores again confirm the lack of reliability (below 85% agreement) of the observations recorded by the teacher of student classroom behavior.

Observer-Mediator Reactivity

In order to determine observer-mediator reactivity effects, three 2x2x2x5 repeated measures analyses of variance were calculated on the frequency of teacher praises, prompts, and criticisms addressed to students in each condition during Interval I and Interval II. In addition, the same analysis was calculated using the total number of teacher verbalizations (praises + prompts + criticisms) as the dependent measure. The independent observer's data were used in all analyses. A significant condition x interval interaction was predicted. It was predicted that the teacher's verbalizations to the students in the experimental condition would increase when she was recording their behavior during Interval II: the number of verbalizations to the students in the experimental and control conditions would not differ significantly during Interval I but would differ significantly in Interval II when the teacher recorded the classroom behavior of the students in the experimental condition.

Praises (see Table 11). The results of the analysis revealed no statistically significant main effect and no significant interactions. The frequency of teacher praises to the students in the experimental condition did not differ from the frequency of praises to the students

in the control condition when the teacher recorded the behavior of the students in the experimental condition during Interval II. Thus, the results failed to confirm the hypothesis of observer-mediator reactivity effects for praises.

Prompts (see Table 12). The analysis revealed a significant condition x interval interaction ($F = 7.93$; $df = 1, 7$; $p < .05$). No main effects and no other interactions were significant.

The means of the condition x interval interaction were compared via a Newman-Keuls test. The interaction is depicted graphically in Figure 3. The results showed that the mean number of prompts from the teacher to the students in the experimental and control conditions did not differ significantly in Interval I and the mean number of prompts to the students in the control condition did not differ significantly from Interval I to Interval II. There was a significant increase in the mean number of prompts from the teacher to the students in the experimental condition from Interval I to Interval II at .05 level. The mean number of prompts for students in the experimental condition was significantly greater than for the students in the control condition during Interval II, at .05 level. The results showed that the teacher did change her verbal behavior to the students in the experimental condition when she was recording their classroom behavior and therefore confirmed the prediction of observer-mediator reactivity effects for prompts.

Criticisms (see Table 13). The analysis indicated a significant referral status main effect ($F = 6.15$; $df = 1, 7$; $p < .05$). The mean

number of criticisms the teacher directed to the referred students selected by the teacher on the basis of a high frequency of off-task behavior was significantly greater than the mean number of criticisms to the nonreferred students, selected by the independent observer, throughout Interval I and Interval II. There were no other significant main effects or interactions. The results for criticisms did not support the hypothesis of observer-mediator reactivity effects since the number of criticisms to the students in the experimental and control conditions did not differ significantly during Interval II. The teacher did not change the frequency of criticisms to the students in the experimental condition when she was recording their classroom behavior during Interval II.

Total Teacher Verbalizations (see Table 14). Praises, prompts, and criticisms were added and the total teacher verbalizations to the students in each condition was used as the dependent measure in the analysis. The results indicated a significant condition x interval interaction ($F = 7.77$; $df = 1, 7$; $p < .05$). No other main effects or interactions were significant.

A Newman-Keuls test was performed on the means of the condition x interval interaction. The interaction is represented graphically in Figure 4. There was no statistically significant difference in the mean number of teacher verbalizations to the students in the experimental and control conditions during Interval I. There was a significant difference in the mean number of teacher verbalizations for the students in

the experimental and control condition during Interval II at .05 level. In addition, there was a significant increase in the mean number of teacher verbalizations to the students in the experimental condition from Interval I to Interval II, at .05 level. There was no significant difference for the mean number of teacher verbalizations to the students in the control condition from Interval I to Interval II. The results support the prediction of observer-mediator reactivity effects since the teacher did change her verbal behavior to the students in the experimental condition concurrently with the recording of their classroom behavior.

CHAPTER IV

DISCUSSION

The purpose of the present investigation was to systematically evaluate the effectiveness of teachers as behavior observers. On the basis of previous studies scrutinizing the methodological problems of using independent observers, it was predicted that teachers would be biased, reactive, and unreliable behavior recorders. In addition, observer-mediator reactivity effects were expected: it was predicted that teachers would alter their own behavior in response to the behavior of the students they were observing.

The results of the study confirmed the predictions of observer reactivity, lack of observer reliability, and observer-mediator reactivity. The teachers were reactive and unreliable observers and observing the behavior of the students resulted in changes in the teachers' verbal interaction behaviors with these students. The prediction of observer bias, however, was not supported.

Observer Bias

The lack of observer bias effects was evidenced by the finding of no significant difference in the data recorded by the teacher as to the percent of appropriate behavior exhibited by the referred and nonreferred students. It had been predicted that the teachers would be biased and record the behavior of the referred students as less appropriate then the nonreferred students since the teachers had personally selected the referred students on the basis of a high frequency of

"off-task" behavior. The results did not, however, support the supposition that the observer's own expectation as to the behavior of the individuals they were observing would result in observer bias effects (Johnson & Bolstad, 1973).

The results of previous investigations of the effects of observer expectations on the data recorded by independent observers have been inconclusive. Several studies reported that experimentally-induced observer expectations resulted in observer bias effects (Kass & O'Leary, 1970; Rosenthal & Fode, 1963; Scott, Burton, & Yarrow, 1967). Several other studies (Kent, O'Leary, Diament, & Dietz, 1973; Skinrud, 1971) failed to confirm these results.

O'Leary, Kent, and Kanowitz (1974) have suggested that the results of previous studies in which observer expectations have resulted in observer bias in the data recorded by independent observers may have been confounded by the experimenter's differential reinforcement of data concordant with experimental hypotheses. They contend that observer expectations alone do not result in observer bias effects. In support of this proposition they have experimentally demonstrated that the data recorded by independent observers may be "shaped" by differential experimenter feedback to conform to experimental hypotheses.

It seems probable that experimenter evaluation of data in the presence of the independent observers may unintentionally contribute to the manifestation of observer bias effects through the differential reinforcement of data consistent with the experimenter's expectations. Furthermore, it seems equally probable that biased verbal reports of

behavior observations may reflect the observer's desires to please the experimenter. If undergraduates are receiving course credit or professional recommendations in exchange for their services as behavior observers, they may deem it more profitable to record and/or report data that is consistent with experimental hypotheses than data that is accurate but inconsistent with the experimenter's expectations.

In the present study, the independent observer intentionally refrained from the discussion of the data recorded by the teacher during the five days that the teachers were recording student classroom behavior. The independent observer answered all teacher references to the recorded data by saying that all the data would be evaluated at the end of the week of baseline observations. In addition to controlling for the differential effects of experimenter feedback, this procedure was employed in order to avoid effecting changes in the teacher's scoring criteria that might have also confounded the measure of teacher-observer reliability. Thus, the lack of observer bias effects in the present study would seem to be supportive of O'Leary et al's contention that observer bias effects reflect differential experimenter reinforcement rather than the observer's expectations of the observee's behavior.

Furthermore, reports from the teachers during the observation period, in addition to the lack of recorded observer bias effects, actually suggested that the systematic observation of student behavior obliviated prior teacher expectations and resulted in a less biased subjective evaluation of the classroom behavior of the observed students. Several teachers commented that they had been unaware of how frequently

the nonreferred student exhibited "off-task" behavior. Likewise, several teachers expressed surprise at how infrequently the referred student exhibited "off-task" behavior.

Observer Reactivity

Observer reactivity effects were demonstrated in the present investigation by changes in the percent of appropriate behavior exhibited by the students when the teachers began to record their classroom behavior. Some students exhibited increases and other students exhibited decreases in appropriate behavior when teacher observations were initiated.

The finding of differential reactivity to teacher observations is concordant with numerous prior reports of individual differences in reaction to being observed (Moos, 1968). White (1973) reported that the age of the child being observed affected the degree of deviant behavior suppressed in the presence of an outside observer. Similarly, Patterson and Harris (1968) found that the observee's base-rate level of family interaction behaviors interacted with the direction of behavior change exhibited in the presence of an outside observer: those individuals exhibiting an initial low level of interaction increased the frequency of their interaction behaviors whereas individuals exhibiting an initial high level of interaction decreased the frequency of their interaction behaviors in the presence of an observer (regression to the mean effect).

Furthermore, Mash and Hedley (1974) have found that in addition to characteristics of the observee, the nature of the individual's past

history of interactions with the observer can affect the direction and magnitude of the behavior change. The presence of an adult who had previously interacted positively with the child resulted in increments in performance whereas the presence of an adult who had previously interacted negatively with the child produced decrements in the child's performance of a simple motor task. Mash and Hedley (1974) conclude that "the presence or absence of observer effect, its magnitude, directionality and persistence would appear to be a function of a complex interaction between observer, performer and situation characteristics." In the present study, the magnitude and directionality of the behavior change of each student was most probably a result of observer and observee characteristics, including the initial base-rate level of appropriate behavior and the quality of the student's prior history of interactions with the teacher.

In addition, it is probable that the changes in the teacher's verbal interaction behaviors that occurred concurrently with the teacher recording the classroom behavior of the students (observer-mediator reactivity) contributed to the changes in the students' behaviors that were observed. In the present study it is possible that increases in appropriate behavior were attributable to contingent teacher attention for appropriate behavior. On the other hand decreases in appropriate behavior may have reflected contingent teacher attention for other more passive or disruptive behaviors.

Several researchers (Foster, Keilitz, & Thomas, 1974; Kubany & Sloggett, 1973; Surratt, Ulrich, & Hawkins, 1969) have suggested that

use of the teacher as the observer of student behavior might provide a less obtrusive, and thus less reactive, measurement of student behavior than the addition of independent observers to the classroom. The results of the present study, however, indicate that teacher observations are also reactive, effecting changes in the students' behavior.

Observer Reliability

In the present study, the inter-observer agreement (reliability) was below 35% agreement, substantially below an acceptable level of reliability. Generally, a minimum inter-observer agreement score of 75-85 percent agreement is required in studies in which naturalistic observation procedures are employed as a means of behavior assessment.

The considerably lower reliability obtained in this investigation than in studies in which independent observers were employed as behavior recorders may have reflected insufficient observer training. Typically, when independent observers are utilized as behavior recorders, actual observation sessions are preceded by a training period. Training sessions ordinarily continue until a predetermined level of reliability is obtained. Extensive training in the use of the observation procedure had been purposively avoided in this study in order to approximate the usual procedure in the classroom environment. If observations by the teacher had been preceded by a training period, it is likely that a higher level of reliability would have resulted.

On the other hand, the reliability score obtained may have actually been inflated by the presence of the independent observer during all

teacher observation sessions. Reid (1970), Romanczyk, Kent, Diament, and O'Leary (1973), and Taplin and Reid (1973) have demonstrated the reactive improvement in reliability when reliability assessment procedures are overt. In the present study the teachers were not explicitly informed that reliability was being assessed. It is likely however that several of the teachers surmised that the independent observer was recording the classroom behavior of the same students. In addition, the presence of the independent observer eliminated the opportunity for the teachers to "fake" their observations by recording on the data sheets without observing the students' behavior. The "faking" of observation data has been reported in studies employing independent observers as behavior recorders (Rosenthal & Lawson, 1964; Verplanck, 1965). The fabrication of data by teachers serving as observers is an event whose probability is most likely well above zero in a busy teacher's classroom.

Furthermore, unlike the independent observer, whose exclusive role in the classroom is the observation and recording of behavioral data, the teacher must assume additional, prepotent educational functions. The teacher's academic responsibilities may interfere with her recording of student behavior, resulting in a low inter-observer reliability score. Simkins (1971) has offered a similar explanation for the lower reliability of data recorded by self-recorders than by trained independent observers.

Observer-Mediator Reactivity

Observer-mediator reactivity effects were evidenced by the finding that the teachers altered their verbal behavior toward the students they were observing concomitantly with the initiation of the teacher observation procedures. Specifically, the teachers increased the frequency of total verbalizations and prompt statements during the time that they were recording student classroom behavior.

The specificity of the observer-mediator reactivity effects to an increase in prompt statements may have been attributable to the observation procedure. The observation procedure required the teacher to glance over at the students whose behavior they were recording 15 times during each observation session regardless of the students' behavior at that time. Since the students the teachers were observing had been purposively selected for the study because they exhibited a high frequency of "off-task" behavior, the teacher was likely to find the students engaged in other than task-related behaviors at each observation. A prompt to begin or continue to work would therefore have been an appropriate teacher verbalization at that time. It would be interesting to examine the differential observer-mediator reactivity effects that might result from the teacher utilizing various recording procedures. For example, an increase in criticism statements might be effected by requesting the teacher to keep a frequency count of a student's inappropriate behavior. Similarly, a frequency count of an appropriate behavior might effect an increase in teacher praise statements.

In the current study, the independent observer was present in the classroom during all sessions in which the teachers recorded student behavior. Furthermore, the independent observer had explicitly cautioned the teachers, prior to the initiation of the teacher observation interval, not to alter their behavior in any manner during the recording of student behavior. Mercatoris and Craighead (1974) have experimentally demonstrated that the presence of an independent observer in the classroom produces changes in teacher behavior. Hursh, Baer, and Rowbury (1974), in addition to confirming these results, found that the direction or specificity of the change in teacher behavior was in concordance with the experimental instructions. Teachers were more likely to carry out the experimenter's procedures when independent observers were present in the classroom than when they were absent. In view of the fact that in the present study teachers had been warned against modifying their behavior during the observation sessions, it seems probable that the observer-mediator reactivity effects that were manifested may actually have been attenuated as a result of the reactive effects of the presence of the independent observer on the classroom behavior of the teacher. It is important to recall, however, that the independent observer was present before and during the interval that teachers recorded student behavior. Thus, the reactive effects of the independent observer's presence were not confounded with the finding of observer-mediator reactivity effects.

Referral Status

In addition to the delineation of the methodological problems in the use of teachers as behavior observers, the present investigation resulted in some specification of the kind of behavior that lead teachers to label certain students as "off-task." Teachers were requested to select two students (referred students) from their classrooms who frequently exhibited "off-task" behaviors. "Off-task" was defined as behavior other than task-related such as daydreaming, talking without permission, doodling on paper and out of seat. The operational definition of "off-task" was purposively broad so as not to bias the teacher toward the selection of students who engaged in predominantly either passive or disruptive behaviors. In addition, two other students (nonreferred students) who frequently displayed "off-task" behavior were selected from each teacher's classroom by an independent observer on the basis of time samples of student classroom behavior.

The results of the study indicated that the referred and nonreferred students in the experimental condition did not differ with respect to the percent of appropriate behavior exhibited. The nonreferred students in the control condition, however, exhibited significantly less appropriate behavior than the referred students in the control condition throughout all experimental sessions. This result demonstrates that even though the teachers did select students exhibiting high frequencies of "off-task" behavior, they did not select other students in their classroom who were exhibiting as high, if not higher rates of "off-task" classroom behavior. This suggests that the referred students

were exhibiting a different pattern of "off-task" behavior than the nonreferred students, which resulted in the teachers labelling these students as "off-task."

The finding that the nonreferred students were significantly more passive than the referred students throughout the experiment suggests that those specific behaviors that lead to referral are behaviors which are disruptive to the teacher or other students in the classroom. It appears that students exhibiting predominantly passive "off-task" behaviors are less likely to be referred than students engaged in more disruptive "off-task" behaviors that may interfere with the procedures of the classroom.

The results of the study also revealed that the frequency of teacher criticism statements to the referred students was significantly greater than to the nonreferred students throughout the experiment. Thus even though the referred students were receiving a higher frequency of verbal criticism statements from the teacher, they were, nonetheless, exhibiting off-task behavior. Taking into consideration that the mean number of teacher verbalizations to each student during each 15-minute observation session was only .53, it is likely that teacher criticism statements functioned as reinforcers, increasing the frequency of behavior that resulted in teacher attention.

Summary

In summary, the results of this study indicate that the use of teachers as behavior observers is subject to the same, plus additional

methodological problems that have been extensively investigated in the use of independent observers. The implications of this finding should be of concern to both researchers who employ teachers as the sole source of data collection and to school psychologists who rely upon the data recorded by teachers to assess baseline rates of student target behaviors and to evaluate treatment effectiveness.

The results of prior research studies and case reports that have relied exclusively upon teachers as data recorders may have been confounded by methodological problems. It is imperative that behavioral researchers be aware of the effects of these observational problems on the quality and utility of their experimental investigations. A lack of observer reliability may add unsystematic error to a study thereby reducing the probability of detecting a significant relationship. In addition, even when significant relationships are delineated, low observer reliability may jeopardize the external validity or generalizability of these findings. Even more importantly, observer reactivity and observer-mediator reactivity may endanger both the internal and external validity of experimental findings.

School psychologists must also be made aware of the problems inherent in the use of teachers as behavior observers. They, too, must guard against the problems of a lack of reliability, observer reactivity and observer-mediator reactivity when drawing conclusions about treatment effectiveness.

In conclusion, some mention should be made of the possible therapeutic effects of observations by the teacher on the behavior of the

observed student. The majority of studies concerning the reactive effects of self-recording have reported substantial therapeutic benefits from instructing individuals to record their own behavior. Self-recording has almost attained the level of an initial treatment technique in the area of self-control (Kanfer, 1970). It seems that the reactive effects of observations by the teacher may likewise occasionally contribute to improvements in behavior. Reports of "baseline cures" or decreases in the frequency of undesirable behaviors during the recording of baseline data demonstrate the possible therapeutic benefits that may accrue from teacher observations of student behaviors (Crowder & Willis, 1972). More research is necessary however before any conclusion concerning the therapeutic usefulness of this technique can be proclaimed, in view of the fact that in the present study an equal number of students exhibited behavioral deteriorations as behavioral improvements in response to the teachers observing their classroom behavior.

BIBLIOGRAPHY

Arsenian, J. Young children in an insecure situation. Journal of Abnormal and Social Psychology, 1943, 38, 225-249.

Ayllon, T. and Azrin, N. H. Reinforcement and instructions with mental patients. Journal of the Experimental Analysis of Behavior, 1964, 7, 327-331.

Bales, D. F. Interaction process analysis. Cambridge: Addison-Wesley, 1950.

Bandura, A. Principles of behavior modification. New York: Holt, Rinehart, and Winston, 1969.

Bechtel, R. B. The study of man: Human movement and architecture. Transaction, 1967, 4, 53-56.

Becker, W. C., Madsen, C.H., Arnold, C. R., and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom problems. Journal of Special Education, 1967, 1, 287-307.

Bersoff, D. N. and Ericson, C. R. A precise and valid measure of behavior and behavior change in the classroom. Journal of School Psychology, 1972, 10, 361-366.

Candland, D.K., Dresdale, L., Leiphart, J., and Johnson, C. Videotape as a replacement for the human observer in studies of nonhuman primate behavior. Behavior research methods and instrumentation, 1972, 4, 24-26.

Callahan, E. J. and Alevizos, P. N. Reactive effects of direct observation of patient behaviors. Paper presented at meeting of the American Psychological Association, Montreal, September, 1973.

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand, McNally and Co., 1963.

Cobb, J. A. and Ray, R. S. Manual for coding discrete behaviors in the school setting. Unpublished manuscript, 1971.

Crowder, J. and Willis, J. Implementing behavior modification in Head Start classes using regular inservice and behavioral inservice approaches. Unpublished manuscript, 1972.

Forehand, R. Teacher recording of deviant behavior: A stimulus for behavior change. Journal of Behavior Therapy and Experimental Psychiatry, 1973, 4, 39-40.

Foster, C., Keilitz, I., and Thomas, D. Data collection by the teacher: Interference or stimulus control. Paper presented at the American Psychological Association Annual Convention, New Orleans, September, 1974.

Grimm, J. A., Parsons, J. A., and Bijou, S. W. A technique for minimizing subject-observer looking interactions in field settings. Journal of Experimental Child Psychology, 1972, 14, 500-505.

Hall, R. V., Christler, C., Cranston, S., and Tucker, B. Teachers and parents as researchers using multiple baseline designs. Journal of Applied Behavior Analysis, 1970, 3, 247-255.

Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, F., and Porcia, E. The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. Journal of Applied Behavior Analysis, 1971, 4, 141-149.

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart, and Winston, 1963.

Hursh, H. B., Baer, D. M., and Rowbury, T. A pilot project to examine whether teachers "turn on" only when observers are present. Paper presented at the Annual Convention of the American Psychological Association, New Orleans, September, 1974.

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 practice. Champaign, Illinois: Research Press, 1973.

Johnson, S. M. and Lobitz, G. Demand characteristics in naturalistic observation. Unpublished manuscript, University of Oregon, 1972.

Kanfer, F. H. Self-monitoring: Methodological limitations and clinical applications. Journal of Consulting and Clinical Psychology, 1970, 35, 148-152.

Kass, R. E. and O'Leary, K. D. The effects of observer bias in field-experimental settings. Paper presented at a Symposium "Behavior Analysis in Education," University of Kansas, Lawrence, Kansas, 1970.

Kent, R., O'Leary, K.D., Diament, G., and Dietz, H. Expectation biases in observational evaluation of therapeutic change. Journal of Consulting and Clinical Psychology, 1974, in press.

Kubany, E. S., Bloch, L. E., and Sloggett, B. B. The good behavior clock: A reinforcement/timeout procedure for reducing disruptive classroom behavior. Journal of Behavior Therapy and Experimental Psychiatry, 1971, 2, 173-179.

Kubany, E. S. and Sloggett, B. B. Coding procedure for teachers. Journal of Applied Behavior Analysis, 1973, 2, 339-343.

Lipinski, D. P. and Nelson, R. O. Problems in the use of naturalistic observation as a means of behavioral assessment. Behavior Therapy, 1974, 5, 341-351.

Mash, E. J. and Hedley, J. Observer effect as a function of prior history of social interaction. Submitted for publication, 1974.

Mash, E. J. and Makohoniuk, G. The effects of prior information and behavioral predictability on observer accuracy. Unpublished manuscript, University of Calgary, 1974.

McAllister, L. W., Stachawiak, J. G., Baer, D. M., and Conderman, L. The application of operant conditioning techniques in a secondary school classroom. Journal of Applied Behavior Analysis, 1969, 2, 227-285.

Mercatoris, M. and Craighead, W. E. Effects of nonparticipant observation on teacher and pupil classroom behavior. Journal of Educational Psychology, 1974, 66, 512-519.

Mischel, W. Personality and Assessment. New York: Wiley, 1968.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. Psychological Review, 1973, 80, 252-283.

Moos, R. H. Behavioral effects of being observed: Reactions to a wireless transmitter. Journal of Consulting and Clinical Psychology, 1968, 32, 383-388.

O'Leary, K. D. The effects of observer bias in field-experimental settings. Unpublished manuscript, State University of New York at Stony Brook, 1973.

O'Leary, K. D. and Kent, R. Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds.) Behavior change: Methodology, concepts, and practices. Champaign, Illinois: Research Press, 1973.

O'Leary, K. D., Kent, R. N., and Kanowitz, J. Shaping data collection congruent with experimental hypotheses. Journal of Applied Behavior Analysis, 1974, in press.

O'Leary, K. D., Romanczyk, R. G., Kass, R. E., Dietz, H., and Santogrossi, D. Procedures for classroom observation of teachers and children. Unpublished manuscript, State University of New York at Stony Brook, 1971.

Patterson, G. and Harris, A. Some methodological considerations for observation procedures. Paper presented at the meeting of the American Psychological Association, San Francisco, 1968.

Patterson, G. R., Ray, R. S., and Shaw, D. A. Direct intervention in families of deviant children. Oregon Research Bulletin, 1969, 8.

Patterson, G. R. and Reid, J. B. Reciprocity and coercion: Two facets of social systems. In C. Neuringer and J. L. Michael (Eds.) Behavior modification in clincal psychology. New York: Appleton-Century-Crofts, 1970.

Polansky, N., Freeman, W., Horowitz, M., Irwin, L., Papanis, N., Rappaport, D., and Whaley, F. Problems of interpersonal relations in research on groups. Human Relations, 1949, 2, 281-291.

Reid, J. B. Reliability assessment of observation data: A possible methodological problem. Child Development, 1970, 41, 1143-1150.

Reid, J. B. and DeMasters, B. The efficacy of the spot-check procedure in maintaining the reliability of data collected by observers in quasi-natural settings: Two pilot studies. Oregon Research Bulletin, 1972, 12.

Repp, A. C., Deitz, D. D., Boles, S. M., Deitz, S. M., and Repp, C. F. Differences among common methods for calculating interobserver agreement in applied behavioral studies. Paper presented at the meeting of the Southeastern Psychological Association, Hollywood, Florida, May, 1974.

Roberts, P. R. and Renjaglia, G. A. The influence of tape recording on counseling. Journal of Counseling Psychology, 1965, 12, 10-16.

Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 1, 175-184.

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

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

Scott, P., Burton, R. V., and Yarrow, M. Social reinforcement under natural conditions. Child Development, 1967, 38, 53-63.

Simkins, L. The reliability of self-recorded behavior. Behavior Therapy, 1971, 2, 598-601.

Skindrud, K. A preliminary evaluation of observer bias in multivariate field studies of social interaction. Unpublished manuscript, University of Oregon, 1971.

Surratt, P. R., Ulrich, R. E., and Hawkins, R. P. An elementary student as a behavioral engineer. Journal of Behavior Analysis, 1969, 2, 85-92.

Taplin, P. S. and Reid, J. B. Effects of instructional set and experimenter influence on observer reliability. Child Development, 1973, 44, 547-554.

Verplanck, N. S. The control and the content in conversation: Reinforcement of statements of opinion. Journal of Abnormal and Social Psychology, 1955, 51, 668-676.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand, McNally and Co., 1966.

White, G. D. Effects of observer presence on mother and child behavior. Paper presented at the Annual meeting of the Western Psychological Association, Anaheim, California, 1973.

Yates, A. J. Behavior therapy. New York: John Wiley & Sons, Inc., 1970.

Appendix A

Observation Sheet Used by Independent Observers for Selection of Nonreferred Students

Subject ________________________ Date ____________

Teacher ________________________ Time ____________

+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-	-	-

Appendix B

Definition of Codes for Student Classroom Behaviors

Symbol	Behavior	Definition
A	On-Task	Engaged in behaviors required for the completion of task assigned by teacher. Includes reading, writing, counting on fingers, etc. that are related to the specifically assigned task.
P	Passive	Engaged in behaviors other than task-related behaviors which are not disruptive to other members of the class. Includes daydreaming, sleeping, drawing pictures, working on other classroom work than the specifically assigned task.
D	Disruptive	Engaged in behaviors other than task-related behaviors that are disruptive to the other members of the class. Includes out of seat, talking, making noises, etc.

Appendix C

Definition of Codes for Teacher Verbal Interaction Behaviors

Symbol	Behavior	Definition	Examples
+	Praise	Verbal responses indicating that the teacher is satisfied or pleased with the academic and/or social behavior exhibited by the student.	"Correct" "Good work" "That's right" "I like the way you are behaving today." "You are working so nicely." "Yes"
0	Prompt	Verbal responses conveying additional information necessary to complete assigned tasks or directing the student's attention toward the task.	"Turn the page" "Which other one is _____" "Finish your work" "Listen carefully" "Work carefully"
-	Criticism	Verbal responses indicating that the teacher is not satisfied or is displeased with the academic and/or social behavior exhibited by the student.	"Wrong" "No" "I'm not happy with your behavior today" "That's not right"

Appendix D

Observation Sheet Used by Independent Observers During Intervals I and II

Name ____________________________ Teacher _______________

Observer __________________________ Date __________________

Time __________________

A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+
P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o
D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-

A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+
P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o
D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-

A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+
P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o
D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-

A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+
P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o
D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-

A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+	A	+
P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o	P	o
D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-	D	-

Appendix E

Instructions for Teachers

Behavior Management Program
Title VI Project
Guilford County Schools

DIRECTIONS FOR RECORDING STUDENT BEHAVIOR

(A sample data sheet is attached to this set of instructions -- look at this data sheet while reading the directions.)

STEP 1: Fill in the starting time.

STEP 2: At the starting time indicated, the observer will set the timer for the number of minutes specified in the first space below the column marked "Time."

STEP 3: When the timer bell rings, glance over at the two students and record their behavior in the appropriate column, at that instant as either:

A (Appropriate): Attending to the assigned task.

P (Passive): Not attending to the assignment but not disturbing others in the class.

D (Disruptive): Disturbing others -- e.g., out of seat, talking without permission or making other noise.

STEP 4: The timer will be reset for the next specified number of minutes.

OBSERVATION DAY 1: Before beginning to record, read the following to the class:

"May I have your attention please? Every once in a while this timer will ring. Do not pay any attention to it. I am doing this to keep track of certain things I want to do. Simply disregard the bell and continue to work."

Appendix E (Continued)

After the first several times the timer rings, certain class members may attend to the bell and even make comments about it. The teacher should ignore all references to the timer. In a short period of time, the class should adapt to the bell, and it will not disrupt the students when they are working.

Data will be collected during one hour of each day for five consecutive school days. Try to record during the same times of each day.

Thank you again for your participation in this project.

Appendix F

Observation Sheet Used by Teachers

Starting Time _____

Time	Student 1	Student 2
2		
5		
7		
4		
6		
3		
1		
3		
1		
4		
7		
5		
6		
4		
2		

Appendix G

Tables and Figures

Table 1

Reliability of Independent Observers for Student and Teacher Behaviors

Behavior	Exact Agreement	Pearson (significance level)
Student	.92	.97 ($p < .001$)
Appropriate (A)	.91	.98 ($p < .001$)
Passive (P)	.91	.95 ($p < .001$)
Disruptive (D)	.94	.99 ($p < .001$)
Teacher	.92	.97 ($p < .001$)
Praise (+)	.93	.96 ($p < .001$)
Prompt (O)	.92	.98 ($p < .001$)
Criticism (-)	.90	.98 ($p < .001$)
Student and Teacher (mean reliability)	.92	.97 ($p < .001$)

Table 2

Observer Bias: Referral Status (2) x Days (5) Repeated Measures Analysis of Variance for Appropriate Student Behavior

Source	df	MS	F
Between Subjects			
Subj. w. groups	7	0.715	
Within Subjects			
Referral Status (R)	1	0.115	0.154
R x Subj. w. groups	7	0.745	
Days (D)	4	0.320	2.464
D x Subj. w. groups	28	0.130	
R x D	4	0.161	0.947
RD x Subj. w. groups	28	0.170	

Table 3

Observer Reactivity: Condition (2) x Referral Status (2) x Interval (2) x Days Within Interval (5) Repeated Measures Analysis of Variance for Appropriate Student Behavior

Source	df	MS	F
Between Subjects			
Condition (C)	1	0.108	0.130
Referral Status (R)	1	1.922	2.313
C x R	1	3.798	4.571*
Subj. w. groups	28	0.831	
Within Subjects			
Interval (I)	1	0.139	0.431
C x I	1	0.130	0.403
R x I	1	0.070	0.217
C x R x I	1	0.003	0.008
I x Subj. w. groups	28	0.323	
Days within Interval (D(I))	8	0.347	1.642
C x D(I)	8	0.055	0.260
R x D(I)	8	0.121	0.572
C x R x D(I)	8	0.346	1.637
D(I) x Subj. w. groups	224	0.212	

*$p < .05$

Table 4

Observer Reactivity: Condition (2) x Referral Status (2) x Interval (2) x Days Within Interval (5) Repeated Measures Analysis of Variance for Passive Student Behavior

Source	df	MS	F
Between Subjects			
Condition (C)	1	0.025	0.030
Referral Status (R)	1	7.092	8.359**
C x R	1	1.051	1.239
Subj. w. groups	28	0.848	
Within Subjects			
Interval (I)	1	0.033	0.203
C x I	1	0.037	0.232
R x I	1	0.139	0.859
C x R x I	1	0.002	0.010
I x Subj. w. groups	28	0.161	
Days within Interval (D(I))	8	0.160	1.067
C x D(I)	8	0.065	0.435
R x D(I)	8	0.113	0.754
C x R x D(I)	8	0.286	1.908
D(I) x Subj. w. groups	224	0.150	

**$p < .01$

Table 5

Observer Reactivity: Condition (2) x Referral Status (2) x Interval (2) x Days Within Interval (5) Repeated Measures Analysis of Variance for Disruptive Student Behavior

Source	df	MS	F
Between Subjects			
Condition (C)	1	1.059	0.934
Referral Status (R)	1	2.788	2.459
C x R	1	0.564	0.497
Subj. w. groups	28	1.134	
Within Subjects			
Interval (I)	1	0.237	0.999
C x I	1	0.005	0.021
R x I	1	0.097	0.408
C x R x I	1	0.016	0.061
I x Subj. w. groups	28	0.237	
Days within Interval (D(I))	8	0.151	0.984
C x D(I)	8	0.050	0.323
R x D(I)	8	0.087	0.565
C x R x D(I)	8	0.063	0.413
D(I) x Subj. w. groups	224	0.153	

Table 6

Change in Classroom Behavior for Each Student From Interval I to Interval II

Condition	Appropriate	Passive	Disruptive
Experimental-Referred			
S_1	-00.2	-00.6	+00.8
S_2	-07.6	+09.6	-02.0
S_3	-27.4	+10.6	+16.8
S_4	-03.6	-02.4	+09.0
S_5	+24.8	-13.0	-11.8
S_6	+24.4	+12.0	-36.4
S_7	-04.6	-00.8	+05.4
S_8	+14.4	-04.6	-10.0
Experimental-Nonreferred			
S_9	-03.6	+05.6	-02.0
S_{10}	+13.4	-13.4	00.0
S_{11}	+08.2	-10.4	+02.2
S_{12}	+28.4	-18.4	-10.0
S_{13}	+03.6	+05.0	-08.6
S_{14}	-22.0	+16.2	+05.8
S_{15}	-27.6	+22.2	+05.4
S_{16}	+10.8	-17.6	+06.8
Control-Referred			
S_{17}	+13.6	-05.6	-08.0
S_{18}	-10.2	+14.8	-04.6
S_{19}	+09.4	-12.2	+02.8
S_{20}	-11.8	+11.6	+00.2
S_{21}	+19.4	-06.2	-03.2
S_{22}	-13.4	+14.0	-00.6
S_{23}	+04.0	+00.2	-04.2
S_{24}	+00.4	+08.2	-08.6

Table 6 (Continued)

Condition	Appropriate	Passive	Disruptive
Control-Nonreferred			
S_{25}	+06.8	-07.2	+00.4
S_{26}	+06.0	+02.8	-08.8
S_{27}	-02.6	-03.8	+06.4
S_{28}	-07.0	+09.4	-02.4
S_{29}	+08.0	-08.6	+00.6
S_{30}	-06.4	-00.2	+06.6
S_{31}	-05.8	+16.8	-11.0
S_{32}	-04.0	-07.0	+11.0

Table 7

Observer Reactivity: Condition (2) x Referral Status (2) Analysis of Variance for the Absolute Change in Appropriate Student Behavior from Interval I to Interval II

Source	df	MS	F
Condition (C)	1	286.801	4.443*
Referral Status (R)	1	19.531	0.303
C x R	1	66.701	1.033
Subj. w. groups (error)	28	64.553	

*$p < .05$

Table 8

Observer Reactivity: Condition (2) x Referral Status (2) Analysis of Variance for the Absolute Change in Passive Student Behavior from Interval I to Interval II

Source	df	MS	F
Condition (C)	1	35.701	1.240
Referral Status (R)	1	45.601	1.584
C x R	1	162.901	5.660*
Subj. w. groups (error)	28	28.781	

*$p < .05$

Table 9

Observer Reactivity: Condition (2) x Referral Status (2) Analysis of Variance for the Absolute Change in Disruptive Student Behavior from Interval I to Interval II

Source	df	MS	F
Condition (C)	1	83.851	2.011
Referral Status (R)	1	46.561	1.116
C x R	1	128.801	3.088
Subj. w. groups (error)	28	41.705	

Table 10

Reliability of Teacher Observations: Teacher-Independent Observer Reliability (Agreement)

Behavior	Exact Agreement	Pearson (significance level)
Student	.43	.29 ($p < .001$)
Appropriate	.58	.40 ($p < .0005$)
Passive	.43	.23 ($p < .0359$)
Disruptive	.27	.25 ($p < .0225$

Table 11

Observer-Mediator Reactivity: Interval (2) x Condition (2) x Referral Status (2) x Days Within Interval (5) Repeated Measures Analysis for Teacher Praises

Source	df	MS	F
Between Subjects			
Subj. w. groups	7	3.332	
Within Subjects			
Interval (I)	1	0.003	0.008
I x Subj. w. groups	7	0.396	
Condition (C)	1	1.378	0.603
C x Subj. w. groups	7	2.285	
Referral Status (R)	1	8.128	3.707
R x Subj. w. groups	7	2.192	
Days within Interval (D(I))	8	0.745	1.371
D(I) x Subj. w. groups	56	0.543	
I x C	1	0.903	4.023
IC x Subj. w. groups	7	0.225	
I x R	1	0.078	0.322
IR x Subj. w. groups	7	0.242	

Table 11 (Continued)

Source	df	MS	F
C x D(I)	8	0.539	0.669
CD(I) x Subj. w. groups	56	0.805	
R x D(I)	8	0.486	1.231
RD(I) x Subj. w. groups	56	0.395	
I x C x R	1	0.028	0.136
ICR x Subj. w. groups	7	0.207	
C x R x D(I)	8	0.411	0.775
CRD(I) x Subj. w. groups	56	0.530	

Table 12

Observer-Mediator Reactivity: Interval (2) x Condition (2) x Referral Status (2) x Days Within Interval (5) Repeated Measures Analysis for Teacher Prompts

Source	df	MS	F
Between Subjects			
Subj. w. groups	7	92.796	
Within Subjects			
Interval (I)	1	12.403	2.997
I x Subj. w. groups	7	4.139	
Condition (C)	1	89.253	3.083
C x Subj. w. groups	7	28.946	
Referral Status (R)	1	81.003	4.738
R x Subj. w. groups	7	17.096	
Days within Interval (D(I))	8	8.430	1.334
D(I) x Subj. w. groups	56	6.319	
I x C	1	28.203	7.939*
IC x Subj. w. groups	7	3.553	
I x R	1	3.403	0.874
IR x Subj. w. groups	7	3.896	

Table 12 (Continued)

Source	df	MS	F
C x R	1	42.778	2.384
CR x Subj. w. groups	7	17.942	
C x D(I)	8	3.814	0.443
CD(I) x Subj. w. groups	56	8.603	
R x D(I)	8	4.617	0.681
RD(I) x Subj. w. groups	56	6.785	
I x C x R	1	0.528	0.100
ICR x Subj. w. groups	7	5.265	
C x R x D(I)	8	5.266	0.712
CRD(I) x Subj. w. groups	56	6.215	

*$p < .05$

Table 13

Observer-Mediator Reactivity: Interval (2) x Condition (2) x Referral Status (2) x Days Within Interval (5) Repeated Measures Analysis for Teacher Criticisms

Source	df	MS	F
Between Subjects			
Subj. w. groups	7	4.407	
Within Subjects			
Interval (I)	1	1.378	1.755
I x Subj. w. groups	7	0.785	
Condition (C)	1	4.753	4.423
C x Subj. w. groups	7	1.075	
Referral Status (R)	1	1.953	6.153*
R x Subj. w. groups	7	0.317	
Days within Interval (D(I))	8	0.978	1.527
D(I) x Subj. w. groups	56	0.641	
I x C	1	1.378	1.340
IC x Subj. w. groups	7	1.028	
I x R	1	0.003	0.003
IR x Subj. w. groups	7	1.010	

Table 13 (Continued)

Source	df	MS	F
C x R	1	0.903	3.568
CR x Subj. w. groups	7	0.253	
C x D(I)	8	0.566	0.730
CD(I) x Subj. w. groups	56	0.774	
R x D(I)	8	0.697	0.954
RD(I) x Subj. w. groups	56	0.731	
I x C x R	1	0.078	0.156
ICR x Subj. w. groups	7	0.499	
C x R x D(I)	8	0.491	0.818
CRD(I) x Subj. w. groups	56	0.600	

*$p < .05$

Table 14

Observer-Mediator Reactivity: Interval (2) x Condition (2) x Referral Status (2) x Days Within Interval (5) Repeated Measures Analysis for Total Teacher Verbalizations

Source	df	MS	F
Between Subjects			
Subj. w. groups	7	177.050	
Within Subjects			
Interval (I)	1	22.050	5.241
I x Subj. w. groups	7	4.207	
Condition (C)	1	162.450	3.242
C x Subj. w. groups	7	50.107	
Referral Status (R)	1	171.113	4.826
R x Subj. w. groups	7	35.455	
Days within Interval (D(I))	8	11.886	1.177
D(I) x Subj. w. groups	56	10.098	
I x C	1	54.450	7.677*
IC x Subj. w. groups	7	7.093	
I x R	1	4.512	0.551
IR x Subj. w. groups	7	8.184	

Table 14 (Continued)

Source	df	MS	F
C x R	1	63.012	2.062
CR x Subj. w. groups	7	30.555	
C x D(I)	8	6.411	0.433
CD(I) x Subj. w. groups	56	14.818	
R x D(I)	8	7.930	0.704
RD(I) x Subj. w. groups	56	11.266	
I x C x R	1	1.012	0.221
ICR x Subj. w. groups	7	4.570	
C x R x D(I)	8	6.504	0.566
CRD(I) x Subj. w. groups	56	11.501	

*$p < .05$

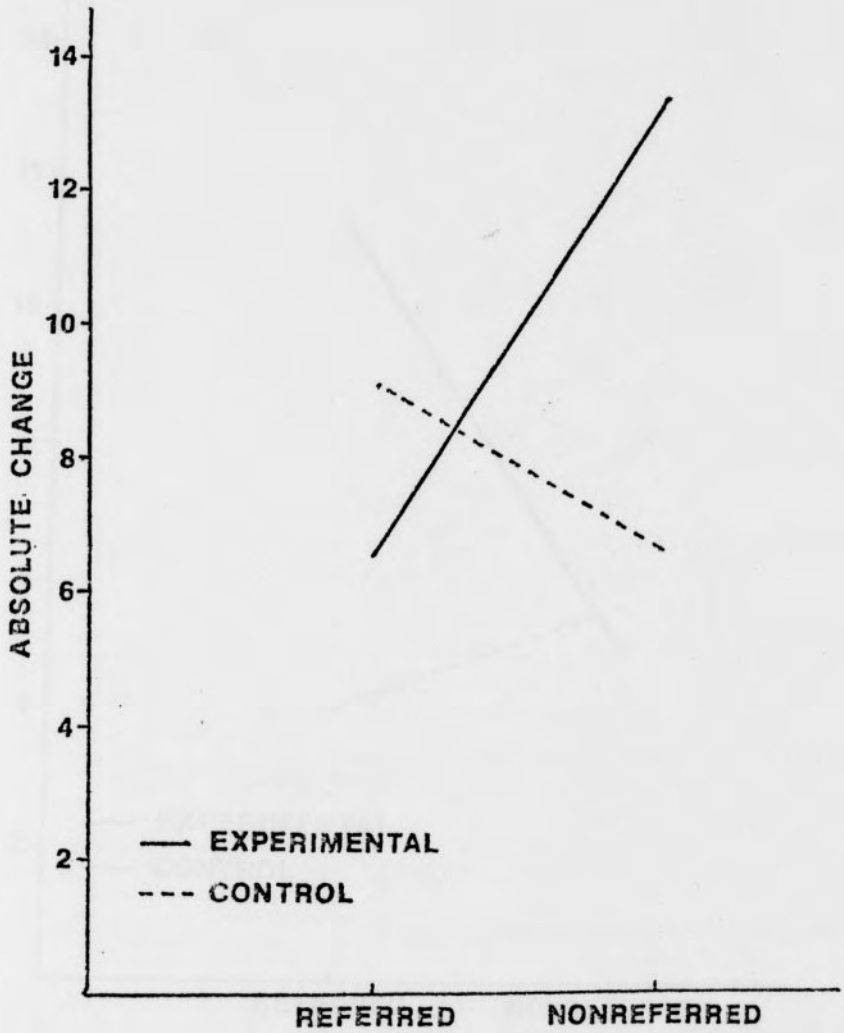

Figure 1. Condition x Referral Status Interaction. Mean absolute change in passive behavior from Interval I to Interval II.

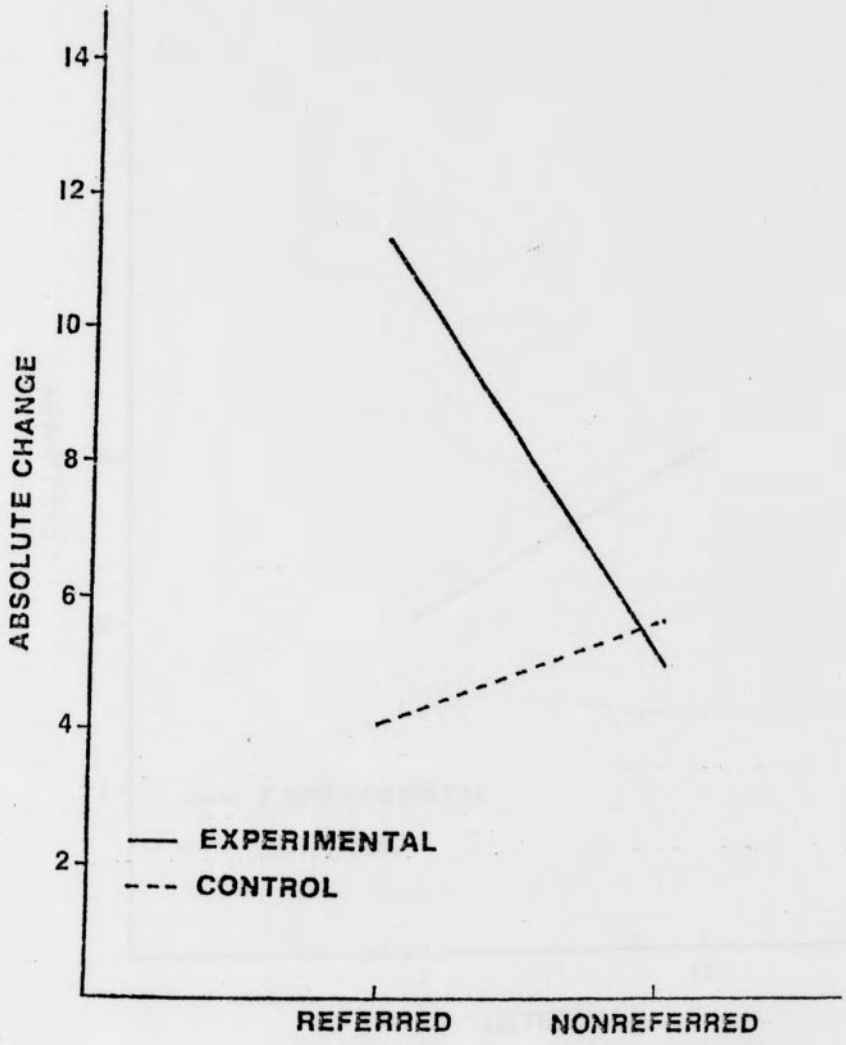

Figure 2. Condition x Referral Status Interaction. Mean absolute change in disruptive behavior from Interval I to Interval II.

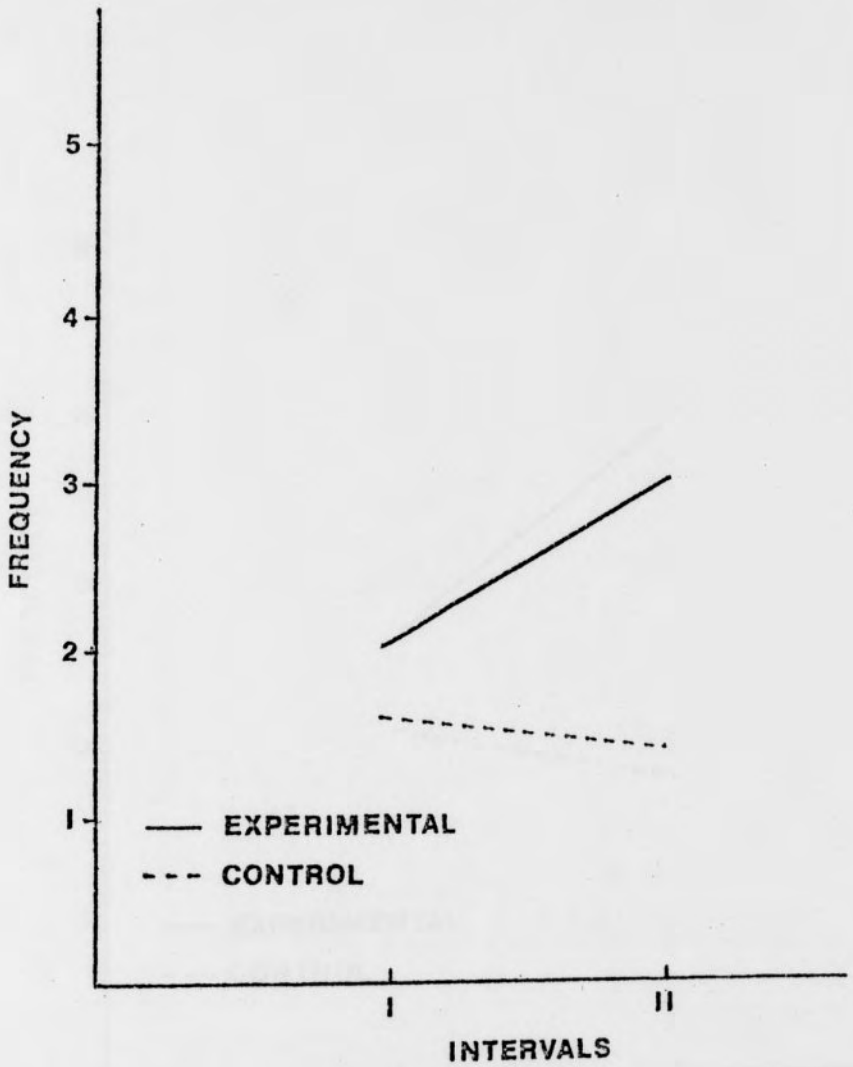

Figure 3. Condition x Interval Interaction. Mean frequency of teacher prompts to students in the experimental and control conditions during Interval I and Interval II.

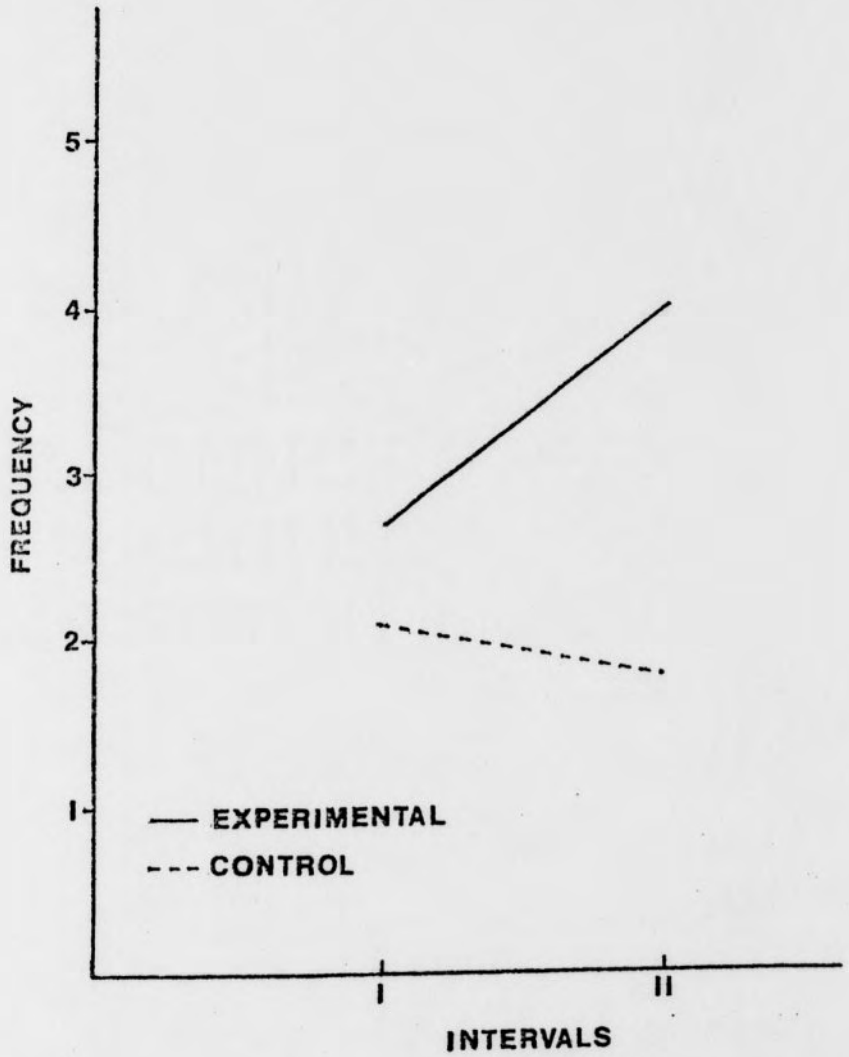

Figure 4. Condition x Interval Interaction. Mean frequency of total teacher verbalizations to students in the experimental and control conditions during Interval I and Interval II.