INFORMATION TO USERS

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

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

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

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

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

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

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR

78-9132

HAY, Linda Rudin, 1951-
METHODOLOGICAL PROBLEMS IN THE USE OF PARTICIPANT OBSERVERS.

The University of North Carolina at Greensboro, Ph.D., 1977
Psychology, experimental

University Microfilms International, Ann Arbor, Michigan 48106

© 1978

LINDA RUDIN HAY

ALL RIGHTS RESERVED

METHODOLOGICAL PROBLEMS IN THE USE OF PARTICIPANT OBSERVERS

by

Linda Rudin Hay

A Dissertation 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
Doctor of Philosophy

Greensboro
1977

Approved by

Rosemary O. Nelson

Dissertation Adviser

APPROVAL PAGE

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

Dissertation Adviser Rosemary O. Nelson

Committee Members Jacquelyn Gaebelein

Arnold Simmel

Robert L. Shalit

William A. Powers

November 9, 1977
Date of Acceptance by Committee

HAY, LINDA RUDIN. Methodological Problems in the Use of Participant Observers. (1977) Directed by: Dr. Rosemery O. Nelson. Pp. 98.

The purpose of the present investigation was to determine if observations by participant observers would effect changes in the frequency of behavior exhibited by observees (observee reactivity) and/or changes in the frequency of behavior exhibited by the observers themselves (observer reactivity). On the basis of prior case studies and research investigations, it was predicted that observations by participant observers would result in both observer reactivity and observee reactivity. In addition, the influence of the valence of the target behavior recorded by the participant observers was investigated to determine if the valence of the behavior would differentially affect the direction or response-specificity of observer reactivity and/or observee reactivity. It was anticipated that the recording of a positively-valenced behavior would produce an increase in response frequency whereas the recording of a negatively-valenced behavior would produce a decrease in response frequency. Furthermore, it was predicted that observation of a positively-valenced behavior would result in an increase in the frequency of positive behaviors exhibited by the observer to the observee whereas observation of a negatively-valenced behavior would result in an increase in the frequency of negative behaviors exhibited by the observer to the observee. The valence of the behavior was similarly expected to affect the level of reliability (inter-observer agreement) of the observations made by the participant observers: higher reliability would be obtained by observers recording a negatively-valenced behavior than by observers recording a positively-valenced behavior.

A multiple baseline design across observees was employed to investigate the relationship between observations by participant observers and changes in the behavior of the observees (observee reactivity) and observers (observer reactivity). Four teachers recorded consecutively the behavior of four of their students. Two teachers recorded appropriate student verbalizations and two teachers recorded inappropriate student verbalizations across all four students. Independent observers also recorded student verbalizations (appropriate and inappropriate) as well as teacher behaviors (positive, negative, and instruction) throughout all phases of the study.

The results of the study substantiated the prediction of observee reactivity in two of the four classrooms. In both of these classrooms, the teachers were recording appropriate student verbalizations (positive valence). In one class, all students exhibited increases in the frequency of appropriate verbalizations with teacher observation. In the other class, two students exhibited increases and two students exhibited decreases in the frequency of appropriate verbalizations when the teacher was observing their behavior. No other changes in student verbalizations were found to be significant in any of the four classrooms. The prediction of observer reactivity was confirmed in only one classroom. One teacher exhibited significant increases in positive and instruction but not negative behavior when she was observing appropriate verbalizations. These observer reactivity effects were in accordance with the predictions concerning the influence of the valence of the target behavior. The valence of the target behavior did not significantly affect the level of reliability obtained by the teachers.

In summary, the results of the present study suggest that in some instances, observations by participant observers may result in changes in the behavior of the observees (observee reactivity) and/or the observers (observer reactivity). The implication of this research should be of concern to researchers employing participant observers in that these methodological confounds may substantially limit the internal and external validity of experimental findings.

ACKNOWLEDGEMENTS

The author would like to express sincere thanks to Dr. Rosemery O. Nelson for her guidance throughout the past five years of graduate training and during the course of this research investigation. Appreciation is also extended to Dr. William Powers for his exceptional assistance with statistical and computer analyses and to Dr. Jacquelyn Gaelelein, Dr. Richard Shull, and Dr. Arnold Rincover for their helpful suggestions and constructive criticisms.

In addition, appreciation is expressed to the principal and teachers of Colfax Elementary School and to the four undergraduate students who served as observers in the present investigation. Special thanks are extended to Ron Boykin for his assistance in observer training and data collection. Finally, the author would like to acknowledge her parents, Irma and Leon Rudin, for without their emotional and financial support, this dissertation would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One of the central assessment techniques used in behavioral research investigations is the direct observation of behavior(s) by independent (nonparticipant) observers. Typically, undergraduate college students or paraprofessionals are trained to use a structured recording procedure and are subsequently situated in the natural environment to record the occurrence of particular responses of interest to the researcher. The extensive reliance on direct observation procedures by behaviorists has been precipitated largely by the recognition of the situational specificity of behavior (Mischel, 1968). Researchers have been concerned that the effects they observed in the laboratory or clinic might not be replicable or generalizable to the extra-laboratory or "real" world environment (Sommer, 1977).

Ironically, this same issue has caused many researchers to question the generalizability of the data collected through direct observations in the natural environment (Johnson & Bolstad, 1973; Kent & Foster, 1977; Lipinski & Nelson, 1974). Researchers have become increasingly worried that the presence of independent observers may itself effect changes in the behavior of the individuals being observed. Inasmuch as the goal of behavioral assessment is to specify objectively what an individual does in response to particular environmental stimuli, there has been concern that the addition of independent observers to the observation environment may limit the generalizability of experimental results.

The effects of the measuring device on the dependent variable, commonly referred to as reactivity (Ciminero, Graham, & Jackson, 1977), has been recognized as a methodological problem in other areas of scientific investigation (e.g., Heisenberg Principle in Physics), as well as psychology, for a number of years. The results of many systematic research investigations concerning the effects of the observers' presence on the behavior of individuals being observed have found observations by independent observers to be reactive. For example, the presence of observers has been shown to effect significant changes in the behavior of nursery school children (Arsenian, 1943), time spent in an exhibit room by museum visitors (Bechtel, 1967), the frequency of interpretive statements made by counselors during therapy sessions (Roberts & Renzaglia, 1965), the frequency of positive verbal behaviors and amount of time mothers played with their children while waiting in a laboratory playroom (Zegiob, Arnold, & Forehand, 1975), the frequency of teacher-student interactions (Mercatoris & Craighead, 1974), teacher compliance with experimenter instructions (Hursh, Baer, & Rowbury, 1974; Sattler & Swoope, 1976), and the reliability of the data recorded by independent observers (Romanczyk, Kent, Diament, & O'Leary, 1973).

Several research studies, however, have reported no significant effects of observer presence. Observer presence was not found to alter significantly the behavior of patients (Callahan & Alevizos, 1974) or staff (Hagen, Craighead, & Paul, 1975) on a psychiatric research ward, the behavior of a discussion group (Bales, 1950), the aggressive behavior of elementary school children (Martin, Gelfand, & Hartmann, 1971), or the

classroom behavior of junior high school students (Nelson, Kapust, & Dorsey, in press) or of retarded students (Mercatoris & Craighead, 1974).

The equivocal findings of these research studies concerning observer presence has led several investigators to hypothesize variables that may account for the discrepant experimental results. Variables that have been suggested to affect differentially the magnitude, direction, or persistence of reactivity include characteristics of the observer (e.g., age, sex, professional status), the observee (e.g., age, sex, test anxiety scores), and the observational setting itself (Mash & Hedley, 1975). In addition, the research paradigm and the nature of the dependent variables selected for observation have probably contributed to the inconsistency of experimental findings (Johnson & Bolstad, 1973).

There is clearly sufficient experimental evidence of observer reactivity to make the presence of observers a potential methodological problem in naturalistic research investigations. Even minimal reactivity effects may substantially limit both the internal and external validity of the results from between-subjects and within-subjects experiments. In between-subjects experiments, unequal reactivity effects across treatment conditions may reduce the internal validity of results by confounding the effects of the independent variable and reactivity effects. The internal validity of within-subjects experiments may be affected if there are changes in reactivity effects over time. On the other hand, reactivity effects may be equal across treatment conditions but limit the external validity or generalizability of experimental findings from both between-subjects and within-subjects experiments.

A study conducted by Dubey, Kent, O'Leary, Broderick, and O'Leary (in press) demonstrates the confounding influence that reactivity effects may have on experimental results. These researchers compared directly recordings of student and teacher behaviors, obtained from behind a one-way mirror when observers were present versus absent from the classroom. Reactivity effects were evaluated for student behavior during an 18-day baseline period and 11 days during which time a classroom token economy was in effect. Nine categories of disruptive student behaviors and 10 categories of teacher behaviors were recorded. Although the majority of student behaviors did not show the presence of the observers to be reactive, the frequency of off-task behavior was significantly altered when observers were present in the classroom. With observers present in the classroom, the frequency of off-task behavior decreased from baseline to treatment. With observers absent from the classroom, the frequency of off-task behavior increased from baseline to treatment. The interaction obtained between the presence of observers and experimental conditions (baseline-token economy) is particularly striking in that statistical analyses revealed significant decreases in the frequency of off-task behavior with the implementation of the token economy only when observers were present in the classroom. The effects of the observers' presence on teacher behavior were evaluated only during the 11 days that the token economy was in effect. Again, the observers' presence was found to be reactive. Teachers made significantly more educational comments when observers were present than when they were absent from the classroom. There was also an increase in a composite measure of teacher behavior

categories. Dubey et al. (in press), however, were unable to replicate this interaction of observer presence and experimental condition although they did report a significant increase in off-task behavior by students when observers were in the classroom, regardless of experimental condition. If only observer present condition had been included, the external validity of the results from this study would have been very questionable. The observer present condition of this study is fairly typical of the experimental conditions comprising behavioral outcome research investigations.

The interaction between observer presence and experimental condition found in this study magnifies the already pervasive concern over the reactive effects of observer presence on the behavior of individuals being observed. The results of this study suggest that researchers may have been overly confident about treatment results obtained from studies in which the data were recorded by nonparticipant observers. The finding that teachers were more compliant with experimental instructions with observers present than when observers were absent (Hursh et al., 1974; Sattler & Swoope, 1976) similarly increases concern over the generalizability of experimental results obtained from studies employing nonparticipant observers.

One solution to the problem of reactivity has been to employ hidden mechanical devices or one-way mirrors in order to record behavior without the individual's awareness (Webb, Campbell, Schwartz, & Sechrest, 1969). Several researchers have used miniature radio transmitters worn by the subjects to collect data in the natural environment (Purcell &

Brady, 1965; Soskin & John, 1963). Gewirtz (1952), Burton (1971), and Bowles and Nelson (1976) have suggested using portable observation booths to record behavior without the observee's awareness.

In the majority of naturalistic research settings, however, the use of hidden or mechanical procedures is impractical and the observer must be visible as he or she records target behaviors. In an attempt to minimize the reactivity engendered by observer presence, investigators have recommended and employed a variety of procedures. Observers have been instructed to "fade into the walls" (Becker, Madsen, Arnold, & Thomas, 1967), to dress and behave in an inconspicuous fashion (Kent & Foster, 1977), to extinguish interactions with the individuals being observed (O'Leary, Romanczyk, Kass, Dietz, & Santogrossi, 1971), or to show "all of the external signs of a piece of furniture" (Peak, 1953). On the other hand, Kinsey, Pomeroy, and Martin (1948) have suggested that the observer become the personal friend of the observee to reduce the influence of the observer's presence. In one study, researchers (Grimm, Parsons, & Bijou, 1972) had observers wear sunglasses so that it would be difficult for the student observees to determine which particular student's behavior was being observed at a given time. They found that children looked at observers wearing sunglasses about half as often as they looked at observers not wearing sunglasses.

Other suggestions to reduce reactivity in the presence of observers have been aimed at the information given to the observee. Kent and Foster (1977) recommend providing observees with a "non-threatening rationale" for the presence of the observers in the setting. This was

the approach Barker and his associates (Barker & Wright, 1951) employed in their studies of the people of Oskaloosa: the people were told that the observers were there specifically for the purpose of learning about the behavior of typical American townspeople. Weick (1968) has proposed that differences in the rationale for observation might account for the discrepancies between studies reporting reactive versus nonreactive effects from observer presence. Johnson and Bolstad (1973) hypothesize that a nonthreatening rationale for being observed might reduce guardedness and anxiety, thereby reducing reactivity.

The problem of reactivity has been ignored by many researchers on the assumption that the effects are short-lived, that is, the initial influence of the observer would gradually decrease as the observer remained in the environment (Deutsch, 1949; Werry & Quay, 1969). A general research guideline has been to have observers present in the observational setting sufficiently long for individuals to "habituate" to their presence prior to actual recording sessions (Jersild & Meigs, 1939; Kent & Foster, 1977; Patterson & Harris, 1968). Data are not yet available, however, about the duration observers must be present for habituation to occur. Thus, this suggestion seems to have been applied on an intuitive nonsystematic basis to date. The few research investigations specifically addressing the "habituation" hypothesis have failed to find the effects of observer presence to diminish over time (Masling & Stern, 1969). The finding that school children continued to look at observers with a high frequency even after the observers had been present frequently in the classroom during a period of 6 months, suggests that exposure to the

observers may not be sufficient to eliminate the reactive effects of observer presence (Grimm et al., 1972). Similarly, Candland, Dresdale, Leiphart, and Johnson (1972) found that the presence of human observers effected changes in the frequencies of certain behaviors exhibited by nonhuman primates following as much as 3 years of contact with human observers. Polansky, Freeman, Horowitz, Irwin, Papania, Rappaport, and Whaley (1949) found, in contrast to habituation, that children observees became increasingly less accepting of the observers' presence at their summer camp over a 3-week period of observations.

Other researchers have attempted to circumvent the issue of reactivity by employing participant observers instead of independent (nonparticipant) observers. Participant observers take a role that is already defined in the social system, thereby providing minimal disturbance to the ongoing behavior in its natural setting. "If the observer takes a role that is already an integral part of the social system, his presence is probably less likely to affect the rest of the system than if he uses a novel role such as psychologist or observer" (Schwitzgebel & Kolb, 1974).

Several strategies of participant observation have been delineated. Most commonly, the observer has been an individual trained to impersonate a member of the social system he is observing. In other studies, however, either ex-members of the group or current group members have been trained to observe the behavior of others in the group. Behavioral researchers have generally preferred to use individuals already present in the environment as participant observers. Individuals who have frequently assumed

the participant observer role in research investigations include teachers (Hall, Christler, Cranston, & Tucker, 1970; Hall, Fox, Willard, Goldsmith, Emerson, Owen, Davis, & Porcia, 1971; Kubany & Sloggett, 1973; McAllister, Stochowiak, Baer, & Conderman, 1969; Osborne, 1969); parents (Harris, 1969; Patterson & Reid, 1970; Zegiob et al., 1975); or peers (Martin et al., 1971; Surratt, Ulrich, & Hawkins, 1969). These individuals, commonly called mediators, usually implement treatment in addition to collecting data.

The results of several case studies and one experimental investigation, however, suggest that observations by mediators may also be reactive, effecting changes in the behavior of the individuals being observed. Reports of "baseline cures" suggest that reactive changes in the observed persons' behaviors may result even when the dependent measures are recorded by participant observers. Crowder and Willis (1972) noted that in several teacher-conducted case studies the frequencies of problem behaviors were markedly diminished when the teachers began baseline observations. Similarly, Forehand (1973) reported a clinically significant reduction in the frequency of spitting behavior following 3 days of baseline observations by teachers. The results of these case studies must be interpreted cautiously, however, because the data indicating behavioral changes were recorded by the teachers themselves without reliability assessments. Thus, it is possible that the data are unreliable and that the students' behavior did not actually change. Further, even if changes in student behaviors were observed, the case study format does

not permit the conclusions that the changes in student behaviors were a result of the experimental procedure (observation) itself.

Hay, Nelson, and Hay (1977) experimentally examined the reactive effects of using teachers as observers of student classroom behavior. Independent observers recorded the behavior of elementary school teachers and students for a 5-day baseline period and for an additional 5-day period during which time each teacher recorded simultaneously the behavior of two of the students in her class. One of the teacher-observed students (referred) had been referred by the teacher for exhibiting a high rate of off-task behavior and the other student (nonreferred) had been selected on the basis of independent observers' recordings of off-task classroom behavior. The teachers were instructed to conduct their classroom activities as usual, and, in addition, to record the behavior of the two students whenever a kitchen timer rang. The kitchen timer rang on the average of every 4 minutes during a 1-hour observation period. The independent observers also recorded the behavior of two control subjects in each classroom, one referred and one nonreferred student, who were not observed by the teacher. The independent observers concomitantly recorded the teachers' verbal interactions with each of the target and control students.

The data recorded by the independent observers suggested that teacher observations were reactive in that those students who were observed by the teacher showed greater changes in behavior than those students who were not observed by the teacher. The direction of these changes in student behavior varied from student to student. Some methodological

problems, however, limit the generalizability of the results from the Hay et al. (1977) study. A bell ring was used as part of the teacher observation procedure to prompt the teacher that it was time to record the students' behavior. Since the bell did not ring during baseline, it is possible that the bell ring itself, rather than the fact that the teachers were recording, may have cued the teachers or students to behave differently. Further, the teachers had no specific training in the observation procedures: the teachers were merely instructed how to record student behavior using the spot-check technique.

It is likely that when participant observers are employed as data collectors, the reactive changes in the observee's behavior are a result of changes in the observer's usual behavior as a member of the group. Both Forehand (1973) and Crowder and Willis (1972) attributed the observed decreases in student behavior to changes in the teachers' responses to the target behaviors. The Hay et al. (1977) study described above also experimentally investigated how observations by teachers affected the teachers' behavior toward the observed students. They found that teachers addressed a significantly greater number of prompts to the students when they were instructed to record student behavior.

Ciminero et al. (1977) also demonstrated experimentally systematic changes in the observers' behavior in a laboratory analogue study. Female college students were assigned to dyads and subsequently appointed to observer and observee roles. Observers were instructed to record the frequency of either leg-kicking and/or face-touching by the observee. The frequencies of these target behaviors exhibited by <u>both</u> observee and

observer were recorded by independent observers situated behind a one-way mirror throughout all experimental phases. The results indicated changes in the observers' behavior (reciprocal reactivity) that were response specific, limited to the particular response being recorded.

The reciprocal reactivity effects reported in this study, however, may have been attributable to experimental procedures other than the fact that the students were recording behavior. The students serving as observers were aware that they too were being observed by the independent observers behind the one-way mirror. Perhaps the changes observed in their own behavior during their observations resulted from their learning which behaviors were of interest to the independent observers. It seems reasonable to hypothesize an increase in response specific reactivity when the observees are aware of which behaviors are being recorded by the observers. This is probably one of the contributing factors in the reactivity frequently reported when individuals self-record their own behaviors (Nelson, 1977). Furthermore, this study was conducted in a laboratory setting rather unrepresentative of the natural environment in which these behaviors occur.

In summary, the results of studies concerning the use of participant observers have been difficult to interpret due to methodological problems. The results of these studies, however, have suggested that the use of participant observers may change not only the behavior of the observees but also the behavior of the observers. In fact, it has been proposed that the changes in observee behavior are attributable to these changes in observer behavior. When a mediator records the behavior of an

individual, he or she is engaging in behaviors that are variant to his or her usual functions in the environment. Therefore, the use of mediators, or participant observers, instead of nonparticipant observers may nonetheless alter the environment in such a way as to effect systematically changes in the behavior of the individuals being observed.

The primary purpose of the present study was to provide a well-controlled experimental investigation, in the natural environment, of the effects of participant observations on the behavior of both observer and observee. For the purposes of this paper, changes in the observees' behavior in response to the observation procedure will be referred to as observee reactivity and changes in the observers' behavior in response to the observation procedure will be called observer reactivity. A within-subjects experimental design was employed and three replications were conducted. The Hay et al. (1977) study employed a group design in which the reactive effects of participant observations were evaluated by comparing a group of teacher-observed students to a group of students who were not observed by the teachers. Although significant changes in both the behavior of the teachers and students were reported, the group comparisons did not demonstrate that the changes in student behavior necessarily correlated with changes in the teacher's behavior toward a particular student. In fact, the group design of the Hay et al. (1977) study did not even permit the conclusion that the changes in student behaviors occurred in those classrooms in which the changes in teacher behaviors were evidenced. In addition to controlling for the methodological confounds in the Hay et al. (1977) study, therefore, the within-subjects

design employed in the present study permitted a closer examination of the correspondence between changes in the behavior of the observers and observees.

Furthermore, a within-subjects experimental design was chosen because it was hypothesized that participant observations would affect the behavior of some observers and not affect the behavior of other observers. The inconsistent results of studies employing independent (nonparticipant) observers have led many researchers to conclude that individuals respond differentially to the presence of independent observers. Even in studies reporting negative results for the effects of observer presence, it has been noted that some individuals appear to react to the presence of the observers. Dubey et al. (in press), for example, stated that the results of their study allowed for the conclusion that the variability associated with observer presence was not greater than that expected by chance in a sample of individuals varying from each other. They admit, however, that their results do not permit the generalization that for any particular child, observations by independent observers will be nonreactive. "It does not seem unlikely that particular children could be identified who would demonstrate substantial increases or decreases in disruptive behavior when observers are present." Similarly, in the present study it was hypothesized that some observers would manifest changes in behavior whereas other observers would not change their behavior while performing the observation task.

Specifically, a multiple baseline design across students was employed to evaluate experimentally the effects of participant observations

on observee and observer behavior individually for each of four teachers. Independent observers recorded the frequency of appropriate and inappropriate student verbalizations of four students in each of four teachers' classrooms. In addition, the independent observers recorded positive, negative, and instruction teacher behaviors addressed to each of the target students. Following a 7-day baseline observation interval, the four teachers, trained in the use of the recording technique, sequentially recorded the behavior of each of the four target students in their class. Each teacher recorded the behavior of one student at a time for 7 days. Thus, each teacher observed student behavior for a total of 28 days. The concurrent recording of student behavior by the teachers and independent observers also permitted the inter-observer agreement (reliability) of teacher observations to be determined.

A second purpose of this study was to evaluate experimentally whether the valence (positive-negative) of the recorded behavior would differentially affect the magnitude or direction of observer reactivity or observee reactivity effects. Two teachers recorded exclusively appropriate student verbalizations (positive valence) and two teachers recorded exclusively inappropriate student verbalizations (negative valence) throughout all experimental phases. Thus, it was possible to make a between-subjects comparison of the effects attributable to the valence of the target behavior.

The results of studies investigating the effects of the valence of the target behavior when individuals are instructed to record their own behavior (self-monitor) have generally found the valence of the target

behavior to be an important factor in determining the direction and magnitude of reactive behavior change. Numerous studies have demonstrated experimentally that the self-monitoring of a positively valenced behavior increases its frequency whereas the self-monitoring of a negatively valenced behavior decreases its frequency. Behaviors with neutral valence have been hypothesized to be less reactive to self-monitoring (Nelson, 1977). Kazdin (1974), for example, manipulated the valence of self-reference statements and found that a positive valence effected increases and a negative valence effected decreases in the frequency of self-reference statements. Similarly, Nelson, Lipinski, and Black (1976) found the social desirability of a behavior to affect the direction of reactive change when adult retardates recorded conversation or face-touching behaviors.

It was predicted that differential reactivity for both observers and observees would be produced by the observation by participant observers of a positively valenced versus a negatively valenced behavior. In the present study, two teachers were instructed to attend to and record appropriate student verbalizations and two teachers were instructed to attend to and record inappropriate student verbalizations. Specifically, it was hypothesized that requesting teachers to keep frequency counts of appropriate student verbalizations would affect an increase in positive teacher verbalizations to the students since the teachers would be more attuned to the occurrence of appropriate student verbalizations. Furthermore, it was predicted that this increase in positive statements would consequently result in an increase in the frequency of appropriate

verbalizations emitted by the observed students. On the other hand, it was hypothesized that requesting teachers to keep frequency counts of inappropriate student verbalizations would effect an increase in negative teacher verbalizations since the teachers would be more attuned to the occurrence of inappropriate verbalizations emitted by the observed students. This increased teacher criticism of inappropriate student verbalizations would result in a decrease in the frequency of inappropriate student verbalizations.

The valence of the target behavior was also hypothesized to alter differentially the reliability of teacher observations. Since teachers are more likely to detect negative behaviors, especially if their occurrence interferes with the behavior of other members of the class, it was predicted that teachers would be more accurate when recording negative than positive student behaviors.

In summary, the following hypotheses were investigated and predictions made:

1. Participant observers would change their behavior toward the observed students when they were performing the observation task. The valence of the recorded behavior would result in differential changes in the observers' behavior. Specifically observers would increase the frequency of positive behavior toward the observed students while recording a behavior with a positive valence and increase the frequency of negative behavior towards the observed students while recording a behavior with a negative valence.

2. Observations by participant observers would result in changes in the observees' behavior with respect to the frequency of the behavior being recorded by the participant observers. The recording of a positively valenced behavior would effect increases in response frequency whereas the recording of a negatively valenced behavior would effect decreases in the frequency of the response.

3. Individual differences between observers would be found. Some observers would evidence consistent changes in their behavior whereas other observers would not exhibit consistent changes in their behavior while performing the observation task.

4. Observers would record a negatively valenced behavior more accurately than a positively valenced behavior.

CHAPTER II

METHOD

Subjects

Teachers. Two first- and two second-grade female Caucasian elementary school teachers were selected from a group of teachers who volunteered to participate in the experiment. Teachers were told that the purpose of the study was to assess how accurately teachers could record student classroom behaviors while conducting their usual classroom activities. The teachers were not informed of the experimental hypotheses concerning observer or observee reactivity until the study was completed. Prior to the initiation of the experiment, each teacher signed a consent form confirming her agreement to participate under these conditions (see Appendix A). Teachers received 2 hours of in-service training credit for their participation.

Students. Sixteen elementary school students, four from each teacher's classroom, participated in the experiment. Twelve students were Caucasian males and four students were Caucasian females. These students were selected on the basis of teacher ratings of five categories of inappropriate classroom verbalizations. Each teacher was asked to complete a short behavior-rating scale for each student in the class (Appendix B). The four students in each classroom receiving the highest total scores on the behavior-rating scale served as target students in that teacher's classroom for the present investigation. Thus, students

were not randomly assigned to teachers. The mean score on the rating scale for subjects in each of the four classrooms was 14, 16.5, 13, and 17.5, respectively. The scores ranged from 9 to 20. The teachers were not informed that a particular student had been selected for the study until they were asked to record that student's behavior.

Experimental Design

A multiple baseline design across students was employed in order to determine whether teacher observations of student behavior were reactive, effecting changes in the behavior of the observed students (observee reactivity) and the behavior of the teacher toward these students (observer reactivity). Independent observers recorded the behavior of the students and teachers for a 7-day baseline period and for 28 additional days during which time each teacher also recorded consecutively the behavior of the four target students in her classroom. The teachers recorded the behavior of one student at a time for 7 school days. The order of observation of these students was determined randomly.

Two teachers recorded only appropriate student verbalizations and two teachers recorded only inappropriate verbalizations throughout the study. This between-subjects comparison allowed for the determination of whether the valence of the target behavior, positive or negative, affected differentially the direction or magnitude of observer reactivity or observee reactivity effects. The effects of this variable on the reliability of teacher observations was also assessed.

Independent Observers

Observer training. Four undergraduate psychology students served as observers. The observers were divided into two observer pairs. Observer pair I recorded student and teacher behaviors in all classes on Mondays, Wednesdays, and Fridays. Observer pair II recorded on Tuesdays and Thursdays. These observers received course credit for their participation.

Independent observers were instructed to wear inconspicuous clothing, sit towards the rear of the classroom as far out of the students' view as possible, and to extinguish interactions initiated by the students. In addition, the teachers were instructed to announce to the students that observers would be present in the room to learn about elementary education procedures. The students were accustomed to being observed because student teachers were often present in the classrooms. These procedures were implemented in order to reduce observee reactivity to the presence of independent observers in the classroom.

Each observer was given written instructions to study that described the observation procedures and behavior codes in detail (see Appendix C). Students from each class were selected at random and their classroom verbalizations were coded by the author and an independent observer for 10-minute practice observation intervals. The author and each observer discussed the codes and procedures following each 10-minute practice observation interval. Practice observations continued until an inter-observer agreement score of at least 85% agreement was obtained between the author and observer for each category on two consecutive observations.

Inter-observer agreement was computed for each code category by dividing the smaller recorded frequency by the larger recorded frequency and multiplying by 100. Observers were kept blind to the experimental hypotheses until the completion of data collection.

Recording procedures. The independent observers recorded student and teacher behavior for 20 minutes a day in each classroom throughout all phases of the study. The independent observers recorded the frequency of appropriate and inappropriate student verbalizations and positive, negative, and instruction teacher behaviors addressed to each target student. A sample data sheet is shown in Appendix D.

Student verbalizations were coded as follows:

1. Appropriate Student Verbalizations: Movements of the mouth that were initiated by a response of the teacher. Examples included answering a question after the teacher had called on the student by name or other gesture, speaking along with the entire class or a group of students following a request by the teacher, and participating in a group oral recitation or singing.

2. Inappropriate Student Verbalizations: Movements of the mouth that were initiated by the student himself or another student in the class. Examples included whispering to a neighboring student, calling out an answer to a question directed to another student or the class as a whole, and interrupting the teacher or another student who was talking.

A new behavior was recorded whenever there was an interval of approximately 5 seconds or greater between verbalizations, or whenever another individual spoke before the target student resumed his (her) verbalizations.

Observers used the wall clock in each room to time the intervals between verbalizations.

Teacher behaviors were coded according to the following three categories:

1. Positive: Verbal or physical responses indicating that the teacher was pleased with the student's behavior. Positive behavior included both responses indicating that the student was correct and responses intended as praise for the student's academic or social classroom behavior. Examples included verbalizations such as "That's good" or "right" and gestures such as a pat on the back or up and down nodding of the head.

2. Negative: Verbal or physical responses indicating that the teacher was displeased with the student's behavior. Negative behavior included both responses indicating that the student was incorrect or responses intended as criticism for the student's academic or social classroom behavior. Examples included verbalizations such as "No, that's wrong" or "I'm disappointed in your work" as well as physical responses such as side-to-side nods of the head.

3. Instruction: Verbal or physical responses conveying information or directing the student's behavior toward a particular task. Examples included instructing, answering questions, giving directions, and calling on a student by name or by pointing.

Inter-observer agreement. A "random-check reliability" procedure (Taplin & Reid, 1973) was used to determine inter-observer agreement. Johnson and Bolstad (1973) report that this reliability assessment

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. Specifically, this procedure required that two observers be recording in the same classroom at the same time. Each observer concurrently recorded the behavior of three of the four target students in each class. The particular three students observed by an observer was randomly determined each day with the stipulation that between the observers all four students were observed each day. The observers were kept unaware of the students whose behavior the other observer was recording. Thus, each day, both observers simultaneously recorded the behavior of two of the four target students in each class. If one observer was absent, the other observer recorded the behavior of all four students in each class on that observation day. This latter circumstance occurred on 3 of the 35 observation days. Thus, inter-observer agreement was determined for 91% of the observation days and 46% of the 20-minute observation intervals.

Spearman correlation coefficients (used for data with nonnormal distributions) were calculated across the daily data recorded by the two observers to determine the inter-observer agreement for the daily frequencies of appropriate and inappropriate verbalizations and for each category of teacher behavior recorded by the independent observers. The coefficients for student appropriate and inappropriate verbalizations were .93 and .88 ($\underline{N}$ = 258), respectively. The agreement coefficients for teacher positive, negative, and instruction behaviors were .83, .62, and .90, respectively.

Inter-observer agreement scores were also computed separately for each of the two observer pairs. Table 1 presents the inter-observer agreement scores for each category of student and teacher behavior as recorded by each observer pair. Observer pair I obtained greater than .85 agreement for both categories of student behavior, and positive and instruction teacher behaviors. Their agreement score for negative teacher behavior, however, was .58. Agreement scores for observer pair II were greater than .75 for all student and teacher behavior recorded.

Teacher Observations

Each teacher was given written instructions describing the observation procedures and behavior code in detail (see Appendix E). The importance of obtaining accurate and objective data was emphasized. Further, the teachers were cautioned against modifying their behavior in any way while they were recording.

Each teacher was given a counter to wear around her neck and on which to record the frequency of student verbalizations. The teachers were instructed to click the counter each time the student being observed exhibited the behavior they were recording. Two teachers were instructed to record appropriate student verbalizations as defined for the independent observers while the other two teachers were instructed to record inappropriate student verbalizations as defined for the independent observers. One first-grade teacher and one second-grade teacher recorded appropriate student verbalizations while the other first- and second-grade teachers recorded inappropriate student verbalizations. The assignment of teachers within each grade level to recording condition was random.

Training sessions for the teachers were conducted for 20 minutes each day. Each teacher practiced recording either the appropriate or inappropriate student verbalizations of one student at a time selected at random from their classes until an inter-observer agreement score of 85% or better was obtained between the author and each teacher on two consecutive 10-minute observations. Inter-observer agreement was computed by dividing the smaller frequency by the larger frequency recorded and multiplying by 100. The author and teacher discussed the recording procedures and behavior definitions following each 10-minute interval.

Procedure

The independent observers recorded student and teacher behaviors in each classroom during a prearranged 20-minute observation interval for 35 consecutive school days. All observations were made during a classroom discussion activity in which the teacher was interacting with the class as a group. Days 1-7 provided baseline frequencies of each target student's classroom verbalizations and the teacher's positive, negative, and instruction behavior to that student. During Days 8-35, the teachers consecutively recorded either the appropriate or inappropriate verbalizations of each of the target students while the independent observers continued to record teacher and student behavior. The teachers recorded the behavior of each student for 7 days. Independent observers signaled the teacher to begin and terminate recording during each observation interval each day.

The teachers were informed that the independent observers were also recording student behavior so that the accuracy of the teachers' recordings

could be determined. The teachers were not informed however that the frequencies of their positive, negative, and instruction behavior were being recorded until the completion of the study. At that time, each teacher was given the option of withdrawing her data from the analysis in order to compensate for failing to obtain prior consent to participate. No teacher utilized this option.

CHAPTER III

RESULTS

Observee Reactivity

Rn statistics (Revusky, 1967) were computed in order to determine whether teacher observations of appropriate or inappropriate student verbalizations effected changes in the rate of occurrence of these student verbalizations. The Rn statistic was suggested by Revusky (1967) specifically for the analysis of data from multiple baseline designs. There are two prerequisites to the use of this analysis procedure: The order of treatment of the individuals must be determined randomly and a minimum of four individuals must receive treatment.

Essentially, the procedure entails viewing the total experiment as a series of subexperiments with one experimental and several control subjects. The individual receiving the experimental manipulation during each phase (interval) of the multiple baseline procedure is considered to be the experimental subject. Following each phase of the study, the individuals are rank ordered with respect to the rate of occurrence (or change in the rate of occurrence) of the target behavior(s). If the manipulation has been effective, then the experimental subject in each subexperiment should rank number 1. If the intervention has not been effective, the rank order should be determined by chance. Following all the subexperiments, the ranks of the experimental subjects from each subexperiment are added and the sum represents the Rn statistic. Each

subexperiment has a probability-generating function of its own which gives the probability that the rank outcome will equal 1: the Rn probability-generating function is determined by multiplying the probability-generating functions of the subexperiments together. Thus, the probability that the Rn statistic was obtained by chance can be determined. A table of values for significance of Rn is available for determining critical values (Revusky, 1967).

The Rn statistic is a nonparametric statistic. As compared with parametric procedures, nonparametric statistics are relatively low powered: there is a higher probability of a Type II error or failure to reject H_0 when in fact it is false. Furthermore, the hypothesis actually tested by the Rn statistic differs from the hypothesis tested with parametric procedures. The Rn statistic tests the hypothesis that all possible rank orderings of the data are equally likely to occur. A significant effect ($p < .05$), therefore, indicates that the rank ordering obtained in the study is likely to occur 5% of the time by chance. Parametric procedures test the hypothesis that the means of the experimental conditions (populations) are equal. A significant effect ($p < .05$) indicates that the obtained difference between the means is likely to be a chance occurrence 5% of the time. Thus, unlike parametric procedures, the Rn test does not take the absolute amount of change between experimental conditions into account in determining significance. Finally, the Rn statistic does not consider the autocorrelation or variance of scores within an experimental condition.

In the present study, Rn statistics were computed for appropriate and inappropriate student verbalizations in each of the four teachers' classrooms using the data recorded by the independent observers. When two observers disagreed on the frequency of a behavior during the 20-minute observation interval, the mean frequency between the two observers was employed in the analyses. In the present study, in order to control for differences in the initial rate of occurrence of the target behaviors across individuals, the absolute difference between the mean frequencies recorded for a subject in each experimental phase to the next phase was employed in the analyses. If the individuals differ prior to the intervention, then the rank orders obtained from each subexperiment may be due to these initial differences rather than the experimental manipulation. Thus, even if the manipulation were effective, it is possible that the effect would be obscured unless change scores were used.

Table 2 presents the mean frequencies across observation intervals of appropriate and inappropriate student verbalizations exhibited by each student in each experimental phase. Table 3 presents the mean change scores between experimental phases that were employed in the Rn analyses. As can be seen in Table 3, the student observed by the teacher during each subexperiment evidenced the greatest change (increase or decrease) in the frequency of appropriate verbalizations in both classes in which the teachers recorded appropriate student verbalizations (Teachers 1 and 2). Summing the ranks for each subject for whom the manipulation was employed in each experimental phase yields Rn = 4 in both classrooms, significant at the $p < .05$ level. In Teacher 1's

classroom, students exhibited increases consistently in the frequency of appropriate student verbalizations when the teacher was observing their behavior. In Teacher 2's classroom, students exhibited either increases or decreases in the frequency of appropriate verbalizations with the experimental manipulation. Figures 1a and 1b graphically depict the frequencies of appropriate student verbalizations for each student in these two classrooms across experimental phase. Teachers 1 and 2 did not observe inappropriate verbalizations. Changes in the frequency of inappropriate verbalizations between experimental phases in these classrooms were not significant. Summing the ranks across subjects yields Rn = 6 and Rn = 8, respectively. Thus, in Teacher 1's and Teacher 2's classrooms, students exhibited significant changes in the frequency of appropriate verbalizations and not inappropriate verbalizations when the teachers were observing appropriate verbalizations.

Table 3 also shows that in the two classes in which the teachers recorded inappropriate student verbalizations (Teachers 3 and 4), the student observed by the teacher in each subexperiment did not consistently evidence the greatest change in the frequency of inappropriate verbalizations. The sums of the ranks across students were Rn = 5 and Rn = 9, respectively. Changes in the frequency of appropriate verbalizations across experimental phases were also not significant in these two classrooms. Summing the ranks across students yields Rn = 6 and Rn = 5 in Teacher 3's and Teacher 4's classrooms, respectively. Thus, in both Teacher 3's and Teacher 4's classrooms, the changes in student behavior with the initiation of teacher observations were not significant.

In summary, observer reactivity effects were evident in two of the four teachers' classrooms. Specifically, when Teachers 1 and 2 were observing appropriate student verbalizations, significant changes in the frequency of the observed behavior were noted. In contrast, when Teachers 3 and 4 were observing inappropriate verbalizations, changes in the observed behavior were not significant.

Observer Reactivity

The previous analyses were concerned with changes in the students' behavior (observee reactivity) with the onset of observations by their respective teachers. Rn statistics (Revusky, 1967) were also computed to determine whether there were changes in the teachers' behavior toward the students (observer reactivity) with the initiation of observations. For each teacher, three Rn statistics were calculated for positive, negative, and instruction behavior, respectively, using the data recorded by the independent observers. Change scores in the mean frequencies of teacher behavior between experimental phases were employed in order to control for differences in the initial rate of occurrence of these behaviors across teachers. Table 4 shows the mean frequencies of positive, negative, and instruction behavior exhibited by each teacher to each student in each experimental phase. Table 5 presents the mean changes across observations in each of these behaviors between experimental phases.

The results of the analyses revealed significant changes in teacher behavior for only one of the four teachers: Teacher 1 evidenced increases in the frequencies of positive and instruction behavior when she was

observing the student's classroom verbalizations. Summing the ranks across experimental subjects yields Rn = 4, significant at the $p < .05$ level for both positive and instruction behavior. Teacher 1 did not exhibit a significant change in the frequency of negative behavior (Rn = 7). Figures 2a and 2b depict the frequencies of positive and instruction behavior exhibited by Teacher 1 to each of the four students across experimental phases. For Teachers 2, 3, and 4, changes in positive, negative, and instruction behavior were not significant. Specifically, for Teacher 2, Rn = 5 for positive behavior, Rn = 6 for negative behavior, and Rn = 7 for instruction behavior. For Teacher 3, Rn = 6 for positive behaviors, Rn = 7 for negative behaviors, and Rn = 6 for instruction behaviors. For Teacher 4, Rn = 6 for positive, negative, and instruction behavior.

In summary, only one of the four teachers evidenced significant changes in teacher behavior with the initiation of observations. Specifically, changes in both positive and instruction teacher behavior were exhibited by Teacher 1 when she was recording appropriate student verbalizations.

Accuracy of Teachers' Recordings (Inter-Observer Agreement Between Teachers and Independent Observers)

Two teachers recorded the frequency of appropriate student verbalizations and two teachers recorded the frequency of inappropriate student verbalizations for 20 minutes a day for 28 school days. Each teacher observed four students consecutively. Thus, the verbalizations of each student were recorded simultaneously by a teacher and the independent

observers for 7 school days. To determine the accuracy of teacher observations, the frequencies of student verbalizations as recorded by the teachers and independent observers were compared. The independent observers' data were used as criteria and considered to be reliable measures of student behaviors because of the high inter-observer agreement between the independent observers (see Inter-Observer Agreement in Chapter II).

Spearman correlation coefficients (used for data with nonnormal distributions) were calculated between the data recorded by each teacher and the independent observers (mean frequency) across all 28 observation days. The agreement scores for Teachers 1, 2, 3, and 4 were .77, .70, .55, and .83, respectively. Thus, the average agreement score across teachers was .71. The average agreement scores across teachers for appropriate and inappropriate student verbalizations were .74 and .69. It had been predicted that teachers would be more accurate in recording a negatively-valenced behavior (inappropriate student verbalizations) than a positively-valenced behavior (appropriate student verbalizations). The teachers' accuracy in recording inappropriate student verbalizations was not significantly higher than their accuracy in recording appropriate student verbalizations; $z = .53$, $p > .10$.

One problem in using correlation coefficients as a measure of observer agreement is that they do not take systematic bias on the part of an observer into account. Spearman correlation coefficients are rank order statistics that reflect "the tendency toward monotonicity and the direction of relationship that appears to exist" (Hays, 1973). Thus, one observer may consistently record fewer or more occurrences of a behavior and this will not be reflected in the magnitude of the correlation.

In the present study, t tests for correlated samples were computed to determine whether the frequencies recorded by the teachers were significantly lower than the frequencies recorded by the independent observers. It was predicted that teachers would record fewer occurrences of the target behavior, in contrast to the independent observers, because they were required to engage in concurrent teaching behaviors while they recorded. Therefore, they were more likely to miss an occurrence of the target response. For all teachers, statistically significant mean frequencies were noted: the mean frequencies of student verbalizations recorded by the teachers were significantly smaller than the mean frequencies recorded by the independent observers. The t values for Teachers 1 and 2 were t (27) = 4.27, $p < .0005$ and t (27) = 2.76, $p < .01$. The t values for Teachers 3 and 4 were t (27) = 2.57, $p < .01$ and t (27) = 4.20, $p < .0005$, respectively. Therefore, the correlation coefficients do not reflect accurately the agreement scores for the teachers because of a systematic bias for the teachers to record lower levels of the behavior than the independent observers. Table 6 shows the mean frequencies of verbalizations recorded by each teacher and the independent observers (mean frequency) for each student.

Figures 3a through 3d depict the frequencies of student verbalizations as recorded by each teacher and the independent observers. It can be seen that the teachers tended to record fewer verbalizations per observation interval than the independent observers. Specifically, the teachers and independent observers recorded the same frequency of student verbalizations during 22% of the observation intervals. The teachers

recorded fewer verbalizations than the independent observers during 65% of the intervals and more verbalizations during 13% of the observation intervals. The valence of the recorded behavior did not appear to affect this distribution. For appropriate student verbalizations, the teachers agreed with the independent observers during 23% of the intervals, recorded fewer verbalizations during 64% of the intervals, and recorded more verbalizations during 13% of the observation intervals. Similarly, for inappropriate student verbalizations, the teachers agreed with the independent observers during 21% of the intervals. The teachers recorded fewer verbalizations during 66% of the intervals, and recorded more verbalizations during 13% of the observation intervals.

An analysis of only those observation intervals during which the teachers and independent observers disagreed on the frequency of occurrence of student verbalizations indicated that 86% of the time the teachers recorded fewer verbalizations than the independent observers. The valence of the recorded behavior did not affect this analysis: for both appropriate and inappropriate student verbalizations, the teachers recorded fewer verbalizations than the independent observers during 84% of the observation intervals.

An additional analysis was calculated in order to determine whether response frequency and teacher reliability scores were inversely related: teachers would be less reliable the higher the rate of occurrence of the target behavior. Spearman correlation coefficients were calculated between the response frequency as recorded by the independent observers and the absolute difference between the frequency recorded by the teacher and

independent observers each observation day. The correlation coefficient for each teacher respectively was .76, .79, .82, and .59. Figure 4 depicts graphically the relationship between teacher reliability and response frequency.

In summary, the level of reliability (inter-observer agreement between the teachers and independent observers) obtained by the teachers in the present study was barely within the criterion (70% - 80% agreement) generally considered as an acceptable level of reliability (Kazdin, 1977). Teachers tended to record fewer instances of the target behavior than the independent observers. Response frequency was found to be related to teacher reliability: the higher the response frequency, the larger the discrepancy between the frequencies recorded by the teachers and independent observers.

Valence of the Target Behavior

It was hypothesized that the valence of the target behavior recorded by the teachers would affect differentially the magnitude and/or direction of changes in the students' behavior (observee reactivity) and changes in the teachers' behavior (observer reactivity). Five analyses of variance and a multivariate analysis of variance were computed on each of the student and teacher behaviors. Each analysis was a 2 [Valence (V)] x 2 [Teachers within Valence (T(V))] x 4 [Students within Teachers within Valence (S(T(S)))] x 2 [Baseline (I_1) - Teacher Observation (I_2)] x 7 [Days (D)] repeated measures analysis of variance. Teachers within valence and students within teachers within valence were considered random variables. Baseline (I_1) was defined as the 7 days immediately

preceding the initiation of observations by the teacher: for Student 1, baseline consisted of Days 1-7; for Student 2, baseline consisted of Days 8-14; for Student 3, baseline included Days 15-21; and for Student 4, baseline was Days 22-28.

The results of the multivariate analysis of variance (Hotelling-Lawley Trace) of the five dependent variables revealed a significant Teachers within Valence (T(V)) effect, F (10, 14) = 3.64, $p < .01$. No other main effects or interactions were significant. The results of the analyses of variance for each of the five dependent variables are presented below.

In each analysis, preliminary tests were conducted and interactions were pooled in order to increase the power of the tests. Thus, if T(V) was nonsignificant, T(V) and S(T(V)) were pooled and the resulting error term was used to test V. Similarly, if IT(V) was nonsignificant, IT(V) and IS(T(V)) were pooled to test D and DV. Finally, if DIV(T) was nonsignificant, DIT(V) and DIS(T(V)) were pooled to test DI and DIV. In order to guard against Type 2 errors, accepting the hypothesis of no interaction when in fact the hypothesis should be rejected, $p < .20$ was employed in all preliminary tests of the model (Winer, 1971).

Appropriate Student Verbalizations

The results of the analysis (Table 7) revealed no significant main effects or interactions. Thus, the valence of the target behavior did not affect significantly the direction and/or magnitude of observee reactivity for appropriate student verbalizations. The mean frequency of appropriate student verbalizations exhibited by the students in each

classroom during baseline and the teacher observation intervals are presented in Table 8.

Inappropriate Student Verbalizations

The results of the analysis (Table 9) indicated significant effects for Days, F (6, 84) = 2.56, $p < .05$, and Teachers within Valence, F (2, 12) = 6.55, $p < .05$. Newman Keuls tests of the mean frequencies of inappropriate student verbalizations per day across teachers and intervals yielded no significant differences between the days. The mean frequencies of inappropriate verbalizations per day across students were 7.93, 4.78, 8.47, 6.28, 4.53, 5.31, and 5.25, respectively. A Newman Keuls test was also computed on the mean frequency of inappropriate verbalizations made by the students in each classroom (Teachers within Valence). The results indicated that the mean frequency of inappropriate verbalizations exhibited by the students in Teacher 3's classroom ($\overline{X}$ = 1.63) differed significantly ($p < .05$) from the mean frequency of inappropriate verbalizations exhibited by the students in Teacher 4's classroom ($\overline{X}$ = 10.27). There was no significant difference between the mean frequencies of inappropriate verbalizations in Teacher 1's classroom ($\overline{X}$ = 4.04) or Teacher 2's classroom ($\overline{X}$ = 8.39). The mean frequencies of inappropriate student verbalizations exhibited by the students in each classroom during baseline (Interval I) and teacher observations (Interval II) are presented in Table 8.

Positive Teacher Behavior

The results of the analysis (Table 10) revealed that there was a significant difference between the teachers within valence, F (2, 12) = 11.32, $p < .05$, as to the frequency of positive behaviors exhibited toward the students across experimental phases. In addition, there was a significant interaction between teachers within valence and intervals, F (2, 12) = 5.28, $p < .05$. Newman Keuls comparisons revealed a significant difference ($p < .01$) in the mean frequency of positive teacher behaviors between Teacher 1 ($\bar{X} = 4.62$) and Teacher 2 ($\bar{X} = 2.80$) and a significant difference ($p < .01$) between Teacher 3 ($\bar{X} = 4.93$) and Teacher 4 ($\bar{X} = 2.34$). Newman Keuls analysis of the interval x teacher within valence interaction revealed a significant difference ($p < .05$) between the frequency of positive teacher behaviors across students exhibited by Teacher 1 in baseline ($\bar{X} = 2.82$) and teacher observations ($\bar{X} = 6.43$). There were no significant differences between teachers with respect to the mean frequency of positive teacher behavior across students during baseline or teacher observations. The mean frequency of positive teacher behavior exhibited by each teacher during each experimental phase is presented in Table 11.

Negative Teacher Behavior

The analysis of variance (Table 12) indicated no significant main effects or interactions indicating that the valence of the target behavior did not affect differentially the frequency of negative behavior exhibited toward the students. The mean frequencies of negative teacher behavior exhibited toward the student during Interval I and Interval II are presented in Table 11.

Instruction Teacher Behavior

The analysis of variance (Table 13) showed a significant difference between teachers within valence, F (2, 12) = 12.90, $p < .05$, and a significant interval x teacher within valence interaction, F (2, 12) = 9.03, $p < .05$. Newman Keuls comparisons revealed a significant difference ($p < .01$) between Teacher 1 ($\bar{X}$ = 10.41) and Teacher 2 ($\bar{X}$ = 6.16) and a significant difference between Teacher 3 ($\bar{X}$ = 7.36) and Teacher 4 ($\bar{X}$ = 3.59) with respect to the mean frequency of instruction teacher behavior exhibited by the teachers across students and experimental phases. Newman Keuls comparisons of the means of the interval x teacher within valence interaction revealed a significant difference ($p < .05$) in the mean frequency of instruction behaviors across students exhibited by Teacher 1 during baseline ($\bar{X}$ = 6.64) and teacher observations ($\bar{X}$ = 14.18). In addition, there was a significant difference ($p < .05$) as to the mean frequency of instruction behavior exhibited by Teacher 1 ($\bar{X}$ = 14.18) and Teacher 2 ($\bar{X}$ = 5.82) during teacher observations. The mean frequency of instruction behavior exhibited by each teacher during each experimental phase is presented in Table 11.

In summary, there was no significant main effect for valence of the target behavior. There was, however, a significant difference between the teachers within valence with respect to the mean frequency of inappropriate student behaviors across their students and across intervals. The students in each classroom differed in the extent to which they engaged in inappropriate behavior. In addition, there were significant differences between the mean frequencies of teacher positive and

instruction behavior across students and intervals exhibited by teachers within valence. Only Teacher 1, who recorded appropriate student verbalizations, however, evidenced a significant change (increase) in the frequencies of positive and instruction behavior from baseline to teacher observation intervals.

CHAPTER IV

DISCUSSION

The primary purpose of the present investigation was to determine systematically if observations by participant observers would effect changes in the frequency of behaviors exhibited by participant observers (observer reactivity) and/or by observees (observee reactivity). On the basis of prior case studies and experimental investigations, it was predicted that observations by participant observers would produce both observer reactivity and observee reactivity. In addition, the influence of the valence of the observed target behavior was also investigated. It was anticipated that the valence of the observed behavior would affect differentially both the direction and response-specificity of observer and observee reactivity effects. The valence of the behavior was similarly expected to affect the level of reliability or the inter-observer agreement of the observations made by the participant observers.

The results of the study substantiated the prediction of observee reactivity in two of the four classrooms. In Teacher 1's classroom, all four of the students exhibited increases in the frequency of appropriate verbalizations when the teacher initiated observations. In Teacher 2's classroom, two students showed increases and two students showed decreases in the frequency of appropriate verbalizations with teacher observations. Both Teacher 1 and Teacher 2 recorded appropriate student verbalizations. In these classrooms, changes in the frequency of inappropriate student

verbalizations did not appear to be associated with observations by the teachers. Teacher 3 and Teacher 4 recorded inappropriate student verbalizations. In these classrooms, systematic changes in neither appropriate nor inappropriate student verbalizations were noted with teacher observations. The results did not confirm the hypotheses that the valence of the target behavior would affect systematically the direction of observee reactivity. Teacher observations of a positively-valenced behavior (appropriate student verbalizations) did not consistently produce increases in response frequency and teacher observation of a negatively-valenced behavior (inappropriate student verbalizations) did not consistently result in decreases in response frequency.

The prediction of observer reactivity was confirmed in only one of the four classrooms. For Teacher 1, the initiation of observations resulted in increases in the frequencies of both her positive and instruction behaviors, but not her negative behaviors. In the other three teachers' classrooms, no significant changes in any of the three observed teacher behaviors (positive, negative, instruction) were revealed with teacher observations. It had also been predicted that the valence of the observed behavior would result in differential changes in the observers' behavior: teacher observations of a behavior with a positive valence would result in increases in the frequency of positive teacher behaviors and teacher observations of a behavior with a negative valence would result in increases in the frequency of negative teacher behaviors. Increases in teachers' instruction behavior were expected in all four classrooms regardless of the valence of the target behavior recorded by the

teacher. Changes in teacher behavior, however, were revealed by only one of the four teachers. Teacher 1 evidenced significant increases in both positive and instruction behaviors, but not her negative behaviors with the initiation of observations. Thus, the direction and response specificity of the significant changes that were detected in Teacher 1's behavior when she was recording student behavior were in accordance with the prediction.

The valence of the target behavior did not affect significantly the level of reliability (inter-observer agreement) obtained by the teachers. Teachers 1 and 2 recorded appropriate student verbalizations (positive valence) and achieved agreement scores (with the data simultaneously recorded by the independent observers) of 77% and 70%. Teachers 3 and 4 recorded inappropriate student verbalizations (negative valence) and obtained agreement scores of 55% and 83%, respectively.

The generalizability of the results of the present study may be limited due to subject selection procedures. The present investigation employed teachers as observers of student behavior in order to test the effect of participant observations on the behavior of the observers and observees. The teacher-pupil relationships described only one type of participant observer-observee model in which the observer assumes an authority role in relation to the observee. The results of the study may not be generalizable to situations in which the observer and observee hold a different type of relationship to one another (e.g., peer observations). Mash and Hedley (1975) have shown that the history of interaction between the observer and observee is an important variable affecting the

direction and magnitude of observer reactivity. Furthermore, the teachers employed were volunteers and may not be representative of teachers who do not volunteer for research investigations. Volunteers are more likely to follow directions and become involved in the experimental procedures. Volunteers are also probably less concerned about the presence of observers in their classroom. Similarly, the students represented a special subset of students in that they were rated by the teachers as the most disruptive students in the class. It is interesting to note, however, that there was generally a low rate of inappropriate verbalizations across students suggesting that the teachers' ratings may not have accurately reflected the students' overt behavior.

In addition, it is important to recall that independent observers were present in all classrooms throughout the study. Although their presence did not affect the internal validity of the study, the external validity or generalizability of results may be limited to situations in which observers are present. The effects of the presence of independent observers in the classroom have been assessed by recording teacher behaviors covertly with observers present and absent from the classroon environment. These studies (Hursh et al., 1974; Mercatoris & Craighead, 1974; Sattler & Swoope, 1976) have concluded that the presence of the independent observers is reactive effecting changes in teacher behavior. It has been noted that teachers are more compliant with experimental procedures when observers are present in the classroom than when they are not present. Additional studies are necessary to determine the generalizability of the results of the present study to other participant

observer-observee relationships and to situations in which independent observers are not present in the classroom.

Observer Reactivity

The results of the Rn analyses indicated that only one teacher evidenced a significant change in the frequency of teacher verbalizations with the onset of teacher observations. Teacher 1 exhibited an increase in the frequency of positive and instruction behavior to the student when she was observing that student's behavior. None of the other three teachers showed significant changes in the frequency of any of the dependent variables with the initiation of observations.

Even though no significant changes in teacher behaviors were detected for three of the four teachers participating in this study, each of these teachers indicated verbally that they thought the observation procedures made them more aware of the behavior of the observed students. Specifically, teacher comments included:

> The main point I've gained during this time is that I was made more aware of my "speaking out" children. I noticed quiet ones who never say or do a thing too.
>
> Although concentrating on the specific actions of an individual student while directing the entire class in a discussion proved challenging, it made me more aware and observant of each student.
>
> I am now more aware of even minor and undisturbing behavior. In the future if some child is having difficulty, this training will help me to pinpoint this child's problem while teaching other students.
>
> This has helped me to watch for minor disturbances in class that might interrupt another student and also to be able to observe the students that are participating in the discussion.

Perhaps there were changes in teacher behaviors that were not reflected by the dependent measures employed in this study. For example, it is likely that the observation procedures effected increases in the frequency with which teachers merely looked or oriented toward the observed students. Further research, including additional or different measures of teacher behavior, seems warranted since the teachers felt unanimously that the observation procedures did alter their classroom behavior in some manner.

Changes in the observer's behavior are potentially problematic when the data are being used to evaluate treatment procedure effectiveness. The results of studies employing participant observers may be confounded by observer reactivity effects. Conclusions concerning treatment effectiveness in these studies may be misleading since the treatment may only be effective if the teacher is also recording student behavior. To separate observer reactivity and treatment effects, participant observers should record the behavior of persons who are not receiving the experimental manipulations in addition to recording the behavior of those persons who are participating in treatment procedures.

The inconsistency of the participant observer reactivity across teachers found in the present study parallels the findings of studies investigating self-monitoring reactivity. The results of studies examining the reactivity of self-monitoring have been inconsistent with some studies demonstrating dramatic reactive effects from self-monitoring alone (Gottman & McFall, 1972; Emmelkamp, 1974; Hay, Hay, & Angle, in press; Herbert & Baer, 1972; Johnson & White, 1971) and other studies failing

to find self-monitoring to be reactive (Berecz, 1972; Hall, 1972; McNamara, 1972; Stollak, 1967). Researchers have already begun to investigate independent variables that predict differentially the magnitude and/or direction of reactive behavior change produced by self-monitoring. Some of the variables that have been demonstrated to affect differentially the magnitude and/or direction of the reactive effects of self-monitoring include: the valence of the behavior (Kazdin, 1974; Nelson, Lipinski, & Black, 1976a); subject motivation for behavior change (Lipinski, Black, Nelson, & Ciminero, 1975; McFall & Hammen, 1971); timing of self-monitoring (Bellack, Rozensky, & Schwartz, 1974; Rozensky, 1974); and the schedule of self-monitoring (Mahoney, Moore, Wade, & Moura, 1973). Similar studies seem indicated to determine the conditions under which observer reactivity occurs or fails to occur. Variables of interest might include characteristics of the observer or observee and the nature of the target behavior or recording procedures.

Observee Reactivity

The results of the Rn analyses for observee reactivity effects indicated significant changes in the observees' behavior with the initiation of observations in two of the four classrooms. Specifically, when Teachers 1 and 2 initiated observations of appropriate student verbalizations, significant changes (increases or decreases) in the frequency of appropriate verbalizations attributable to the observation procedure were noted. Changes in the frequency of inappropriate verbalizations in these classrooms were not significant. When Teachers 3 and 4 initiated observations of inappropriate student verbalizations, no significant changes in

the frequencies of appropriate or inappropriate verbalizations were detected. Thus, the valence of the target behavior may have produced differential effects in that changes in student behavior were only revealed in those classes in which the teachers recorded a behavior with a positive valence.

The specificity of changes in the frequency of student verbalizations to those classrooms in which the teachers recorded appropriate responses may also have been attributable in part to the recording procedure itself. Teachers were instructed to record either appropriate or inappropriate student verbalizations by clicking a golf counter. Appropriate student verbalizations were defined as movements of the mouth that were initiated by a response from the teacher. Therefore, appropriate student verbalizations required a teacher-student interaction. Inappropriate student verbalizations were defined as movements of the mouth that were initiated by the student himself or another student in the class and therefore did not necessitate a teacher-student interaction. This interaction may have made it more likely for the observee to realize that the teacher was observing his (her) behavior when the teacher was recording appropriate than inappropriate verbalizations.

Although changes in student behaviors resulting from teacher observations were confined to those classes in which the teachers observed appropriate student verbalizations, the pattern of change in student behavior was not consistent between classes. In Teacher 1's classroom, all students evidenced increases in the frequency of appropriate verbalizations whereas in Teacher 2's classroom both increases and decreases in

appropriate verbalizations were evidenced. Hay et al. (1977) reported similar findings of differential reactivity to teacher observations: some students demonstrated increases and other students exhibited decreases in the percentage of appropriate behavior with the initiation of observations by the teachers.

Visual inspection of the daily frequency of appropriate verbalizations (Figures 1a and 1b) exhibited by the students in Teacher 1's and Teacher 2's classrooms shows that the changes in student verbalizations with teacher observations in Teacher 2's classroom were less dramatic and less consistent than in Teacher 1's classroom. The data for Teacher 1 (Figure 1a) is unambiguous and the statistical analyses (Rn statistic) supports the visual interpretation of the data as significant. The data in Figure 1b, however, is ambiguous. Although the Rn analysis was significant, visual inspection of the data does not suggest a clear-cut effect. Since the Rn statistic is a nonparametric procedure which does not take all of the information from the data (e.g., variability, autocorrelation) into account, the author has little confidence in the replicability of the significant results obtained in Teacher 2's classroom and warns the reader to interpret these results with caution.

One explanation for the differential findings in the two classrooms where teachers recorded appropriate verbalizations may pertain to the manner in which the teachers employed the recording procedures. It was noted by the independent observers that Teacher 1 tended to record by clicking the counter immediately following each occurrence of the target response. In Teacher 1's classroom, the students seemed to be aware of the new

procedure and several students asked the teacher what the clicker was used for. In contrast, Teacher 2 did not always record immediately and often waited for several responses to occur before clicking the counter multiple times. This difference in procedure may have resulted in the students in Teacher 1's classroom having a more easily discernible contingency between the appropriate response and the counter click than the students in Teacher 2's classroom. The students in Teacher 1's classroom received feedback on a continuous reinforcement schedule whereas the students in Teacher 2's classroom received feedback on a variable ratio (VR) schedule. Furthermore, there was a smaller delay between the responses and feedback (counter click) in Teacher 1's classroom than in Teacher 2's classroom.

In Teacher 1's classroom, the observation procedure appears to have functioned as a treatment procedure for increasing appropriate student verbalizations. Forehand (1973) and Crowder and Willis (1972) similarly reported therapeutic changes in student behaviors with the initiation of observations by the teacher. Forehand (1972) described a case study in which observations by a teacher and teacher aide resulted in a clinically significant decrease in spitting behavior in 3 days. Crowder and Willis (1972) noted desirable decreases in several problematic behaviors following the initiation of teacher observations. Crowder and Willis (1972) referred to these successful cases as "baseline cures."

It is likely that the therapeutic changes in student behavior in the present study, as well as these case reports, resulted from systematic changes in the way the teacher responded to the behavior she was observing.

Forehand (1973) attributed the changes in spitting behavior to the teacher's spontaneous initiation of an extinction procedure. He hypothesized that the observation procedure made the teacher more cognizant of the antecedent and consequent conditions maintaining the problematic behavior. Crowder and Willis (1972) attributed the changes in problematic student behaviors to changes in the teachers' responses to the target behaviors when they were counting the behaviors. In the present study, changes in student behavior may have been attributable to changes in teacher behaviors. Teacher 1 exhibited systematic increases in both positive and instruction behaviors (observer reactivity) with the initiation of observations.

Although Teacher 2 did not evidence consistent changes in teacher behavior across students, it is likely that she exhibited different changes in her behavior while observing different students. This may also have accounted for the finding that in Teacher 2's classroom two students exhibited increases and two students exhibited decreases in appropriate verbalizations with teacher observations. For example, Teacher 2 evidenced increases from baseline in the frequency of positive behavior while observing the students who exhibited increases in appropriate student verbalizations and decreases from baseline in the frequency of positive behavior while observing the students who demonstrated decreases in appropriate verbalizations with teacher observations. These changes, however, were not statistically significant.

In summary, these studies suggest that observations by participant observers, when reactive, may not necessarily result in therapeutic

changes in the frequency of the recorded responses. The direction of the behavior change is likely to be dependent on whether or not the teacher has the skills to spontaneously design and implement a successful treatment program and changes her behavior accordingly. Crowder and Willis (1972) note that the "baseline cures" obtained in several teacher-conducted case studies probably resulted from the teachers using specific treatment methods they had learned in a behavior modification course during the baseline phase.

Many researchers have assumed that observations by participant observers would produce less reactivity than observations by nonparticipant observers. It was thought that the addition of individuals to the observees' environment was the primary factor producing reactivity. Since participant observers were already a part of the observees' environment, it was assumed that their use as observers would not be reactive. The results of the present investigation and the Hay et al. (1977) study, however, question this assumption since the use of participant observers still produced reactivity in some classrooms. Thus, although the use of participant observers may eliminate the addition of individuals to the observees' environment, the use of participant observers may still alter the environment in such a way as to effect changes in the behavior of the individuals being observed.

Furthermore, the results of studies concerning the reactivity of nonparticipant observers have been equivocal. While several studies have found the presence of nonparticipant observers to affect the behavior of the individuals being observed (Arsenian, 1942; Bechtel, 1967; Mercatoris & Craighead, 1974; Roberts & Renzaglia, 1965; Sattler & Swoope, 1976),

several recent studies (Dubey et al., 1977; Nelson et al., in press; Weinrott, Garrett, & Todd, 1977) have found observer presence to produce only minimal or no changes in behavior. Thus, in some situations the addition of nonparticipant observers to the environment may not result in changes in the observees' behavior. In fact, it may be incorrect to assume that observations by nonparticipant observers are more reactive than observations by participant observers. Future research investigations should compare directly the reactive effects of participant versus nonparticipant observations as well as attempt to identify variables that predict when observations by either participant or nonparticipant will be reactive. Such variables might include characteristics of the observers, observees, or recording procedures themselves.

Until procedures to guard against reactivity are identified, researchers employing participant or nonparticipant observers should employ control procedures whenever possible. In between-subjects studies, the same observer should observe all groups and an observation-alone control group should be included in the design. In single subject research, the same observers should also record data throughout all experimental phases. In either design, results should be interpreted carefully since the results may not be generalizable to situations in which no observations or different observation procedures are employed.

Observer Reliability (Teacher-Independent Observer Agreement)

The reliability scores obtained by comparing the daily frequencies recorded by the teachers and independent observers ranged from 55% to 83% agreement. The average teacher reliability score was 71% agreement.

Thus, teachers bordered on an acceptable level of reliability since a minimum inter-observer agreement score between 70-80% agreement is usually required in studies employing naturalistic observation procedures for data collection (Kazdin, 1977).

It is difficult to compare the reliabilities obtained by the teachers in the present study to the reliability scores reported in other research investigations employing teachers as observers due to differences in observer training procedures, the observation procedures and the methods used to calculate reliability. The reliability coefficients obtained in the present study were slightly lower than the reliability scores reported by other research investigators employing teachers as observers. Hall et al. (1971) reported inter-observer agreement scores between 85% and 100% when teachers recorded frequency counts of particular target behaviors. Osborne (1969) reported 100% inter-observer agreement in the recording of the frequency of students' out-of-seat behavior. In contrast, the Hay et al. (1977) study reported teacher reliability below 35% agreement, substantially below an acceptable level of agreement. The Hay et al. (1977) study differed from the present investigation, however, in that the teachers were merely instructed and not explicitly trained in the recording procedures. In the present study, teachers were trained and required to meet an agreement criterion of 85% agreement on two consecutive 10-minute observation intervals before actual observation sessions were initiated. As has been evidenced in the training of independent observers, instructions per se are usually not sufficient to produce reliable observations. In addition, the teachers

in the present study and the studies cited above employed a frequency count recording procedure whereas in the Hay et al. (1977) study, a spot-check observation procedure was utilized.

The studies described above also differed from the present study with respect to the methods used to calculate reliability. The method used to calculate reliability has been shown to have a large effect on the resultant reliability score. The teachers in the Hall et al. (1971) study kept frequency counts and reliability was calculated between the data recorded by a teacher and the experimenter by dividing the smaller of the two scores obtained for a session by the larger, and multiplying by 100. This method of determining reliability has been criticized (Hartmann, 1977) as being heavily dependent on the rate of occurrence of the target behavior: higher rates of occurrence result in higher percentage agreement scores. The teachers in the study conducted by Osborne (1969) also kept frequency counts but reliability was determined by the exact agreement method (agreements/(agreements + disagreements) x 100). This procedure is more conservative than the procedure employed by Hall et al. (1971) in that reliability not only requires similar total frequency scores but also requires that each occurrence of the behavior be recorded by the observers at the same time. Hay et al. (1977) also employed the exact agreement method to determine the reliability of occurrence-nonoccurrence data obtained with a time-sampling procedure. Reliability was calculated for occurrence data only. This method of determining reliability may lead to overestimations of agreement if the probability of agreement by change is not taken into account (kappa).

In the present study, a correlation coefficient was calculated to determine the reliability of data recorded by the teachers. This method has the advantage of permitting the subsequent identification of systematic errors (observer bias) by calculating a *t* test of the difference between correlated scores. For teachers in the present study, a systematic bias was noted as teachers recorded fewer occurrences of the target behavior than the independent observers. A correlation may be misleading, however, if observer errors are correlated.

The reliability scores obtained in the study may have been inflated by the presence of the independent observers in the classroom during all teacher observation sessions. Reid (1970), Romanczyk et al. (1973), and Taplin and Reid (1973) have demonstrated reactive improvements in reliability when reliability assessment procedures are overt. Similarly, research studies have found the reliability of self-recorded data to be significantly better during overt reliability assessment than when reliability assessments were made covertly (Nelson, 1977). As Wiggins has noted:

> It makes a certain amount of sense. We all work a little bit harder when the boss is around. Our opinions are perhaps a shade closer to the boss's in his presence. And we are more accurate in describing events that have been observed by others than we are in recounting exploits that cannot be verified.

In the present study, the teachers were very aware that the independent observers were concurrently recording student verbalizations. The teachers had been informed explicitly that the purpose of the study was to determine how accurately they could record student behavior while conducting their usual classroom activities.

In addition, the presence of the independent observers may have improved reliability by eliminating the opportunity for the teachers to "fake" their observations by recording on data sheets without actually observing student behavior. The "faking" of observation data has been reported in studies employing independent observers as behavior recorders (Azrin, Holz, Ulrich, & Goldiamond, 1961; Rosenthal & Lawson, 1964; Verplanck, 1955). With teachers, it has been observed that compliance to experimental instructions is enhanced when independent observers are present in the classroom environment (Hursh et al., 1974; Sattler & Swoope, 1976).

The data indicate that the teachers consistently recorded lower frequencies of the target behavior than the independent observers. Specifically, teachers recorded fewer verbalizations than the independent observers during 65% of the observation intervals. Similarly, an analysis of only those observation intervals in which the teachers' and independent observers' data disagreed revealed 86% of the discrepancies to be the result of omission errors on the part of the teachers. Furthermore, the higher the response frequency (as recorded by the independent observers) the greater the discrepancy between the teachers and independent observers. These findings are not surprising since unlike the independent observers, whose exclusive function in the classroom is the recording of student behaviors, the teacher must engage in numerous other prepotent educational behaviors.

Simkins (1971) has suggested that the low inter-observer reliability scores obtained when individuals record their own behavior might similarly

be attributable to the incompatibility of other concurrent behaviors and the self-recording response. Epstein, Webster, and Miller (1975) experimentally investigated the effects of requiring an individual to perform an operant response concurrent to the self-monitoring response. The results indicated an increase in self-monitoring errors when the operant response task was initiated. A subsequent study (Epstein, Miller, & Webster, 1976) found efficiency in self-monitoring to decrease as the effort and/or vigilance required to perform the concurrent operant task increased, supporting the inverse relationship between the level of reliability and the performance of competing behaviors.

The results of the present study and other studies employing teachers as observers suggest that teachers may be reliable observers. Variables that appear to influence the reliability of observations recorded by teachers include the observation recording procedures, observer training procedures, and the conditions under which observations are made. To maximize reliability when teachers are employed as data collectors, teachers should be instructed in relatively simple recording procedures and should be asked to record during classroom activities that require them to engage in few competing teacher behaviors. Further research is needed to systematically evaluate the influence of these factors and other variables on the reliability of observations by participant observers.

Summary

In summary, the results of the present study suggest that in some instances observations by participant observers may result in changes in the observees' behaviors as well as changes in the observers' behaviors

toward the observees. Although the reactive effects of observations were evident in only two of the four teachers' classrooms, these changes, particularly in one class, were substantial enough to have potentially confounded the results of an experiment employing the teacher exclusively as data collector. Studies employing nonparticipant observers have also found that in some cases observations may be reactive whereas in other cases reactivity is not evident. It appears that variables other than the observer's status as a member of the observees' environment play an important role in determining whether observations are reactive. Until these variables are identified, the results of the present study indicate that researchers should be aware of these methodological problems when using participant observers and should employ control procedures (observation-only condition) whenever possible.

Similarly, although the teachers in the present study as a group bordered on an acceptable level of reliability, there were large individual differences between classes. This suggests that the level of reliability, as well as reactivity, may be determined by a number of factors. Researchers have already begun to investigate variables affecting the level of reliability obtained by independent observers and self-recorders. Likewise, further research is necessary to evaluate the influence of variables on the reliability of observations by participant observers.

BIBLIOGRAPHY

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

Azrin, N., Holz, W., Ulrich, R., & Goldiamond, I. The control of conversation through reinforcement. Journal of the Experimental Analysis of Behavior, 1961, 4, 25-30.

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

Barker, R. G., & Wright, H. F. One boy's day. New York: Harper & Row, 1951.

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., & Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. Journal of Special Education, 1967, 1, 287-307.

Bellack, A. S., Rozensky, R., & Schwartz, J. A comparison of two forms of self-monitoring in behavioral weight reduction programs. Behavior Therapy, 1974, 5, 523-530.

Berecz, J. Modification of smoking behavior through self-administered punishment of imagined behavior: A new approach to aversive therapy. Journal of Consulting and Clinical Psychology, 1972, 38, 244-250.

Bowles, P. E., & Nelson, R. O. Training teachers as mediators: Efficacy of a workshop versus the bug-in-the-ear technique. Journal of School Psychology, 1976, 14, 15-26.

Burton, R. V. An inexpensive and portable means for one-way observation. Child Development, 1971, 42, 959-962.

Callahan, E. J., & Alevizos, P. N. Reactive effects of direct observation of patient behaviors. Paper presented at the meeting of the American Psychological Association, Montreal, September, 1974.

Candland, D. K., Dresdale, L., Leiphart, J., & 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.

Ciminero, A. R., Graham, L. E., & Jackson, J. L. Reciprocal reactivity: Response-specific changes in independent observers. Behavior Therapy, 1977, 8, 48-56.

Crowder, J., & Willis, J. Implementing behavior modification in Head Start classes using regular inservice and behavioral inservice approaches. Unpublished manuscript, 1972. (Available from Jefferson County Department of Health, Birmingham, Alabama)

Deutsch, M. An experimental study of the effects of cooperation and competition upon group processes. Human Relations, 1949, 2, 199-231.

Dubey, D. R., Kent, R. N., O'Leary, S. G., Broderick, J. E., & O'Leary, K. D. Reactions of children and teachers to classroom observers: A series of controlled investigations. Behavior Therapy, in press.

Emmelkamp, P. M. G. Self observation versus flooding in the treatment of agorophobia. Behaviour Research and Therapy, 1974, 12, 229-237.

Epstein, L. H., Miller, P. M., & Webster, J. S. The effects of reinforcing concurrent behavior in self-monitoring. Behavior Therapy, 1976, 7, 89-95.

Epstein, L. H., & Webster, J. S., & Miller, P. M. Accuracy and controlling effects of self-monitoring as a function of concurrent responding and reinforcement. Behavior Therapy, 1975, 6, 654-666.

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

Gewirtz, J. L. Plans for construction of a portable one-way observation booth. Child Development, 1952, 23, 307-314.

Gottman, J. M., & McFall, R. M. Self-monitoring effects in a program for potential high school dropouts: A time series analysis. Journal of Consulting and Clinical Psychology, 1972, 39, 273-281.

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

Hagen, R. L., Craighead, W. E., & Paul, G. L. Staff reactivity to evaluative behavioral observations. Behavior Therapy, 1975, 6, 201-205.

Hall, R. V., Christler, C., Cranston, S., & 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., & Porcia, E. The teacher as observer and experimenter in the modification of disputing and talking-out behavior. Journal of Applied Behavior Analysis, 1971, 4, 141-149.

Hall, S. M. Self-control and therapist control in the behavioral treatment of overweight women. Behaviour Research and Therapy, 1972, 10, 59-68.

Harris, A. Observer effect on family interaction. Unpublished doctoral dissertation, University of Oregon, 1969.

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

Hay, L. R., Hay, W. M., & Angle, H. V. The reactivity of self-recording: A case report of a drug abuser. Behavior Therapy, in press.

Hay, L. R., Nelson, R. O., & Hay, W. M. Some methodological problems in the use of teachers as observers. Journal of Applied Behavior Analysis, 1977, 10, 345-348.

Hays, W. L. Statistics. New York: Holt, Rinehart, & Winston, 1973.

Herbert, E. W., & Baer, D. M. Training parents as behavior modifiers: Self-recording of contingent attention. Journal of Applied Behavior Analysis, 1972, 5, 139-149.

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

Jersild, A. T., & Meigs, M. F. Direct observation as a research method. Review of Educational Research, 1939, 40, 472-482.

Johnson, S. M., & Bolstad, D. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), Behavior change: Methodology, concepts, and practice. Champaign, Ill: Research Press, 1973.

Johnson, S. M., & White, G. Self-observation as an agent of behavioral change. Behavior Therapy, 1971, 2, 488-497.

Kazdin, A. E. Reactive self-monitoring: The effects of response desirability, goal setting, and feedback. Journal of Consulting and Clinical Psychology, 1974, 42, 704-716.

Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABCs of reliability. Journal of Applied Behavior Analysis, 1977, 10, 141-150.

Kent, R. N., & Foster, S. L. Direct observational procedures: Methodological issues in naturalistic settings. In A. Ciminero, K. Calhoun, & H. E. Adams (Eds.), Handbook for behavioral assessment. New York: John Wiley & Sons, Inc., 1977.

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. Sexual behavior in the human male. Philadelphia: Saunders, 1948.

Kubany, E. S., & Sloggett, B. B. A coding procedure for teachers. Journal of Applied Behavior Analysis, 1973, 6, 339-344.

Lipinski, D. P., Black, J. L., Nelson, R. O., & Ciminero, A. R. Influences of motivational variables on the reactivity and reliability of self-recording. Journal of Consulting and Clinical Psychology, 1975, 43, 637-646.

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

Mahoney, M. J., Moore, B. S., Wade, T. C., & Moura, N. G. M. The effects of continuous and intermittent self-monitoring on academic behavior. Journal of Consulting and Clinical Psychology, 1973, 40, 404-407.

Martin, M. F., Gelfand, D. M., & Hartmann, D. P. Effects of adult and peer observers on boys' and girls' responses to an aggressive model. Child Development, 1971, 42, 1271-1275.

Mash, E. J., & Hedley, J. Effect of observer as a function of prior history of social interaction. Perceptual and Motor Skills, 1975, 40, 659-669.

Masling, J., & Stern, G. Effect of the observer in the classroom. Journal of Educational Psychology, 1969, 60, 351-354.

McAllister, L. W., Stachowiak, J. G., Baer, D. M., & Conderman, L. The application of operant conditioning techniques in a secondary school classroom. Journal of Applied Behavior Analysis, 1969, 2, 227-235.

McFall, R. M., & Hammen, C. L. Motivation, structure and self-monitoring: The role of nonspecific factors in smoking reduction. Journal of Consulting and Clinical Psychology, 1971, 37, 80-86.

McNamara, J. R. The use of self-monitoring techniques to treat nail biting. Behaviour Research and Therapy, 1972, 10, 193-194.

Mercatoris, M., & 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.

Nelson, R. O. Methodological issues in assessment via self-monitoring. In J. D. Cone & R. P. Hawkins (Eds.), Behavioral assessment: New directions in clinical psychology. New York: Brunner-Mazel, 1977.

Nelson, R. O., Kapust, J. A., & Dorsey, B. L. Minimal reactivity of overt classroom observations on student and teacher behavior. Behavior Therapy, in press.

Nelson, R. O. Lipinski, D. P., & Black, J. L. The reactivity of adult retardates' self-monitoring: A comparison among behaviors of different valences, and a comparison with token reinforcement. The Psychological Record, 1976, 26, 189-201. (a)

Nelson, R. O., Lipinski, D. P., & Black, J. L. The relative reactivity of external observations, and self-monitoring. Behavior Therapy, 1976, 7, 314-321. (b)

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

Osborne, J. G. Free time as a reinforcer in the management of classroom behavior. Journal of Applied Behavior Analysis, 1969, 2, 113-118.

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

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

Peak, H. Problems of objective observation. In L. Festinger & D. Katz (Eds.), Research methods in the behavioral sciences. New York: Dryden, 1953.

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

Purcell, K., & Brady, K. Adaptation to the invasion of privacy: Monitoring behavior with a miniature radio transmitter. Merrill-Palmer Quarterly, 1965, 12, 242-254.

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

Revusky, S. H. Some statistical treatments compatible with individual organism methodology. Journal of the Experimental Analysis of Behavior, 1967, 10, 319-330.

Roberts, P. R., & Renzaglia, 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., & O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 1, 175-184.

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

Rozensky, R. H. The effect of timing of self-monitoring behavior on reducing cigarette consumption. Journal of Behavior Therapy and Experimental Psychiatry, 1974, 5, 301-303.

Sattler, H. E., & Swoope, K. F. Teachers as token dispensers: Effect of an observer. Psychology in the Schools, 1976, 13, 97-100.

Schwitzgebel, R. K., & Kolb, D. A. Changing human behavior: Principles of planned intervention. New York: McGraw-Hill, 1974.

Simkins, L. The reliability of self-recorded behaviors. Behavior Therapy, 1971, 2, 83-87.

Sommer, R. Toward a psychology of natural behavior. APA Monitor, 1977, 8, 1.

Soskin, W. F., & John V. The study of spontaneous talk. In R. Barker (Ed.), The stream of behavior. New York: Appleton-Century-Crofts, 1963.

Stollak, G. E. Weight loss obtained under different experimental procedures. Psychotherapy: Theory, Research, and Practice, 1967, 4, 61-64.

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

Taplin, P. S., & 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., & Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally, 1969.

Weick, K. E. Systematic observational methods. In G. Lindsley, & E. Aransen (Eds.), The handbook of social psychology (2nd ed., Vol. 2), Reading, Mass.: Addison-Wesley, 1968.

Weinrott, M. R., Garrett, B., & Todd, N. The influence of observer presence on classroom behavior. Manuscript submitted for publication, 1977.

Werry, J. S., & Quary, H. C. Observing the classroom behavior of elementary school children. Exceptional Children, 1969, 35, 461-476.

Wiggins, J. S. Personality and prediction: Principles of personality assessment. Reading, Pa.: Addison-Wesley, 1973.

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

Zegiob, L. E., Arnold, S., & Forehand, R. An examination of observer effects in parent-child interactions. Child Development, 1975, 46, 509-512.

Appendix A

Teacher Consent Form

I _____ agree to participate in this research project. I have been informed that I will be asked to record student classroom behaviors for 20 minutes a day for 8 weeks and that observers from the University of North Carolina will be present in the classroom during that time. I have been told that the purpose of the study is to determine how well teachers can record student behaviors while continuing to conduct their usual classroom activities and that I will receive additional information concerning the purpose of the project at the completion of the experiment.

_____
Teacher's signature

_____
Dr. Rosemery Nelson

Appendix B

Behavior Rating Scale

Student's Name: Date:

Teacher: Grade:

Instructions: Rate the student named above on each of the behaviors listed below:

0	1	2	3	4
Never	Infrequently	Sometimes	Frequently	Very frequently

Rating

_____ 1. Student talks to other students when the class has been told to work quietly and alone.

_____ 2. Student interrupts the teacher while she is talking.

_____ 3. Student interrupts other students while they are talking.

_____ 4. Student blurts out the answer to questions directed to the entire class.

_____ 5. Student blurts out answers to questions that have been directed to other students.

Appendix C

Directions to Observers

1. Please BE PROMPT and dressed appropriately for visiting an elementary school on each observation day.

2. Enter the classroom quietly and sit in as inconspicuous a spot as possible so long as you can still see the students you are supposed to be observing. Do not initiate conversation with the students. If a student approaches you while you are observing say "I cannot talk with you while I am working." The teacher will also tell the students not to talk with you while you are observing.

3. You will be recording in each teacher's classrooms for 20 minutes a day for approximately 10 weeks. Total observation time each day is 2 hours. A second observer will be observing in the same classroom with you at all times in order to check the reliability of the observations. The two observers should be seated a good distance from one another and the observations should not be discussed. Make sure you begin and terminate recording at the same time.

4. Each observer will record the behavior of the teacher and <u>three</u> students simultaneously for the entire 20-minute observation period. The data sheet will specify the students you are to observe each day. You will be observing different students on different days.

5. RECORDING PROCEDURE:

   Each pair of observers will be given a stopwatch in order to insure that the recording interval is <u>exactly</u> 20 minutes each day.

   Record every appropriate and inappropriate verbalization (see behavior codes) made by each of the students you are observing by making a mark in the correct column on the data sheet.

   Also record every positive, negative, or instruction behavior (see behavior codes) the teacher makes toward any of the three students you are observing by making a mark in the correct column. Record <u>only</u> those teacher behaviors addressed to the students whose verbalizations you are recording!

Appendix C (Continued)

	Student		Teacher		
	Appropriate	Inappropriate	Positive	Negative	Instruction
Johnny		1		1	
Linda	1		1		
Bill		1	1	1	

Example 1:

Johnny is talking to the student next to him and the teacher tells Johnny to "Stop talking."

Example 2:

Linda answers a question after the teacher calls on Linda by name. The teacher says "Good."

Example 3:

Bill blurts out the answer to a question the teacher addressed to the class as a whole. The teacher says "Right, but you should not call out without my calling on you."

Appendix D

Recording Form

Observer _____ Class _____ Date _____ Teacher Date _____

NAME	STUDENT BEHAVIOR		TEACHER BEHAVIOR		
	Appropriate	Inappropriate	Positive	Negative	Instruction
CODE DEFINITIONS	Movements of the student's mouth which are initiated by the teacher. Examples: 1. Answering a question after the teacher has called the student by name or gesture. 2. Participating in a group recitation or singing.	Movements of the student's mouth which are initiated by the student himself or another student. Examples: 1. Whispering to a neighboring student. 2. Blurting out the answer to a question addressed to the class.	Verbal or physical responses indicating that the teacher is pleased with the student's academic or social behavior. Examples: 1. "Right." 2. "I like the way you are behaving today." 3. "Good."	Verbal or physical responses indicating that the teacher is displeased with the student's academic or social behavior. Examples: 1. "No." 2. "That's wrong." 3. "I don't like your behavior at all today. 4. "Stop talking."	Verbal or physical responses conveying information or directing the student behavior toward a particular task. Examples: 1. Instructing and giving directions. 2. Calling on a student by name or by pointing. 3. "Pay attention."

Appendix E

Directions For Teachers

A. Inappropriate Classroom Verbalizations

1. You will be given a counter for the duration of the study. This is to be worn around your neck during the 20-minute observation period each day.

2. During the observation period, the class is to engage in a class discussion activity.

3. The observers from UNC-G will tell you when to begin and terminate recording. They will also tell you which student you are to observe for each day.

4. Procedures: Each time the student you are observing makes an inappropriate classroom verbalization, press the clicker on the counter that you are wearing.

5. Definition: inappropriate classroom verbalizations -- any movement of the student's mouth which is initiated by the student himself or another student in the class.
   Examples -- whispering to a neighboring student;
   student blurts out answer to a question without being called on to respond by the teacher;
   student talks to himself or makes verbal noises

6. You will be asked to record the behavior of students in your class for a total of approximately 30 days.

Thank you for your cooperation!

B. Appropriate Classroom Verbalizations

1. You will be given a counter for the duration of the study. This is to be worn around your neck during the 20-minute observation period each day.

2. During the observation period, the class is to engage in a class discussion activity.

3. The observers from UNC-G will tell you when to begin and terminate recording. They will also tell you which student you are to observe for each day.

Appendix E (Continued)

4. **Procedures:** Each time the student you are observing makes an <u>appropriate classroom verbalization</u>, press the clicker on the counter that you are wearing.

5. **Definition:** appropriate classroom verbalizations -- any movement of the student's mouth which is initiated by the teacher.
   Examples -- student answers a question after being called on by name;
   student participation in a group recitation or sing;
   student answers a question along with the rest of the class

6. You will be asked to record the behavior of students in your class for a total of approximately 30 days.

Thank you for your cooperation!

Appendix F

Tables and Figures

Table 1

Inter-Observer Agreement for Each Category of Student and Teacher Behavior[a]

Behavior	Observers Pair I	Pair II
Student		
Appropriate Verbalizations	.93*	.93*
Inappropriate Verbalizations	.89*	.86*
Teacher		
Positive	.85*	.77*
Negative	.58*	.75*
Instruction	.93*	.83*

[a]As calculated by a Spearman correlation coefficient.

*$p < .001$.

Table 2

Mean Frequency Per Experimental Phase of Appropriate and Inappropriate Student Verbalizations Exhibited by Each Student as Recorded by the Independent Observers

	Appropriate Student Verbalizations					Inappropriate Student Verbalizations				
	Int I	Int II	Int III	Int IV	Int V	Int I	Int II	Int III	Int IV	Int V
T_1:										
S_1	3.71	13.71				1.86	0.71			
S_2	5.71	5.14	15.43			8.57	8.06	6.57		
S_3	8.86	4.00	5.00	11.14		0.57	2.29	2.57	5.14	
S_4	3.57	8.14	6.71	10.43	11.14	4.57	3.00	3.14	3.00	4.86
T_2:										
S_1	12.43	6.29				8.57	6.71			
S_2	7.00	5.43	7.29			6.71	3.71	1.57		
S_3	7.43	8.14	7.71	6.57		11.71	14.71	7.00	9.29	
S_4	8.71	5.71	6.43	5.57	9.00	9.57	12.86	12.14	13.86	19.00
T_3:										
S_1	5.86	7.29				2.00	1.43			
S_2	5.14	8.14	10.71			0.29	0.14	0.43		
S_3	6.86	5.29	4.57	5.71		0.86	0.86	0.86	5.00	
S_4	5.86	5.71	6.14	5.43	10.29	0.86	0.86	2.43	2.14	1.00
T_4:										
S_1	24.14	4.57				6.00	5.00			
S_2	17.43	5.14	3.57			5.86	10.57	13.00		
S_3	22.43	4.86	3.86	4.71		11.57	16.57	9.43	21.29	
S_4	23.73	4.14	2.86	2.29	3.00	11.00	7.86	10.71	9.71	7.29

Table 3

Absolute Change in the Frequencies of Appropriate and Inappropriate Student Verbalizations Between Experimental Phases For Each Student

	Appropriate Verbalizations					Inappropriate Verbalizations				
	I-II	II-III	III-IV	IV-V	Rn	I-II	II-III	III-IV	IV-V	Rn
T1:										
S_2	10.00					1.15				
S_2	0.57	10.29			4*	0.57	1.43			6
S_3	4.86	1.00	6.14			1.72	0.28	2.57		
S_4	4.57	1.43	3.72	0.71		1.57	0.14	0.14	1.86	
T2:										
S_1	6.14					1.86				
S_2	1.57	1.86			4*	3.00	2.14			8
S_3	0.71	0.43	1.14			3.00	7.14	2.29		
S_4	3.00	0.72	0.86	3.43		3.29	0.72	1.72	5.14	
T3:										
S_1	1.43					0.57				
S_2	3.00	2.57			6	0.15	0.29			5
S_3	1.57	0.72	1.14			0.00	0.00	4.14		
S_4	0.16	0.43	0.71	4.86		0.00	1.57	0.29	1.14	
T4:										
S_1	19.57					1.00				
S_2	12.29	1.57			5	4.71	2.43			9
S_3	17.57	1.00	0.85			5.00	7.14	11.86		
S_4	19.61	1.28	0.57	0.71		3.14	2.85	1.00	2.42	

*$p < .05$.

Table 4

Mean Frequency Per Experimental Phase-of Positive, Negative, and Instruction Teacher Behaviors Exhibited To Each Student As Recorded By the Independent Observers

	POSITIVE					NEGATIVE					INSTRUCTION				
	Int. I	Int. II	Int. III	Int. IV	Int. V	Int. I	Int. II	Int. III	Int. IV	Int. V	Int. I	Int. II	Int. III	Int. IV	Int. V
T_1:															
S_1	2.29	6.14				0.71	0.43				4.29	13.14			
S_2	1.86	2.00	7.43			0.71	0.43	0.86			6.86	5.14	17.29		
S_3	4.00	3.00	2.29	5.43		0.71	0.00	1.29	1.00		9.71	4.57	4.71	12.43	
S_4	1.14	3.00	3.86	4.71	6.29	0.29	0.86	0.86	0.86	2.00	5.00	7.29	8.14	12.43	13.86
T_2:															
S_1	3.57	2.43				0.57	1.57				6.57	7.29			
S_2	1.86	1.29	3.29			0.57	0.29	0.29			6.43	4.71	5.71		
S_3	2.71	4.29	3.71	2.28		0.71	1.57	0.29	0.14		5.29	9.29	9.71	4.86	
S_4	2.71	2.00	1.86	1.71	4.00	0.71	1.00	0.43	0.43	0.43	6.57	4.00	3.71	5.14	5.57
T_3:															
S_1	3.29	4.86				0.86	0.43				5.29	6.86			
S_2	3.29	3.86	8.71			0.57	0.43	0.57			4.57	7.29	10.43		
S_3	3.71	1.29	3.29	4.00		0.86	0.29	0.29	0.29		5.86	2.29	5.29	6.71	
S_4	3.29	3.14	5.00	3.29	8.14	1.00	0.57	0.14	0.86	0.29	5.43	5.57	6.71	6.29	10.71
T_4:															
S_1	1.29	2.43				0.29	0.43				1.57	3.71			
S_2	2.43	3.43	4.29			0.71	1.00	1.71			3.00	5.43	5.71		
S_3	1.00	2.57	2.29	2.71		0.43	0.57	0.29	0.43		2.43	3.71	2.86	3.71	
S_4	2.29	1.86	1.43	0.86	1.43	0.86	1.26	0.71	0.57	0.29	5.57	3.57	4.00	2.71	2.86

Table 5

Absolute Change in the Frequencies of Positive, Negative, and Instruction Teacher Behaviors

Between Experimental Phases For Each Student

	POSITIVE					NEGATIVE					INSTRUCTION				
	Int. I-II	Int. II-III	Int. III-IV	Int. IV-V	Rn	Int. I-II	Int. II-III	Int. III-IV	Int. IV-V	Rn	Int. B-I	Int. I-II	Int. II-III	Int. III-IV	Rn
T_1:															
S_1	3.85					0.28					8.85				
S_2	0.14	5.43			4*	0.28	0.43			7^{NS}	1.72	12.15			4*
S_3	1.00	0.71	3.14			0.71	1.29	0.29			5.14	0.14	7.72		
S_4	1.86	0.86	0.85	1.58		0.57	0.00	0.00	1.14		2.29	0.85	4.29	1.43	
T_2:															
S_1	1.14					1.00					0.72				
S_2	0.57	2.00			5^{NS}	0.28	0.00			6^{NS}	1.72	1.00			7^{NS}
S_3	1.58	0.58	1.43			0.86	1.28	0.15			4.00	0.42	4.85		
S_4	0.71	0.14	0.69	2.29		0.29	0.57	0.00	0.00		2.57	0.29	1.43	0.43	
T_3:															
S_1	1.57					0.43					1.57				
S_2	0.57	4.85			6^{NS}	0.14	0.14			7^{NS}	2.72	3.14			6^{NS}
S_3	2.42	2.00	0.71			0.63	0.00	0.00			3.57	3.00	1.42		
S_4	0.15	1.86	1.71	4.85		0.43	0.33	0.72	0.63		0.43	1.14	0.42	4.42	
T_4:															
S_1	1.14					0.14					2.14				
S_2	1.00	0.86			6^{NS}	0.29	0.71			6^{NS}	2.43	2.43			6^{NS}
S_3	1.57	0.28	0.42			0.14	0.28	0.14			1.28	0.85	0.85		
S_4	0.43	0.43	0.57	0.57		0.40	0.55	0.14	0.28		2.00	0.43	1.29	0.15	

*$p < .05$.

Table 6

Frequencies of Student Verbalizations as Recorded By the Teachers and Independent Observers

		Appropriate Verbalizations	
		Teacher	Ind. Observer
Teacher I	S_1	11.00	15.29
	S_2	11.14	15.43
	S_3	9.43	11.14
	S_4	10.14	11.14
Teacher II	S_1	5.43	6.28
	S_2	4.29	7.29
	S_3	5.57	6.58
	S_4	7.43	9.00
		Inappropriate Verbalizations	
		Teacher	Ind. Observer
Teacher III	S_1	0.29	1.43
	S_2	0.71	0.43
	S_3	2.29	5.00
	S_4	0.14	1.00
Teacher IV	S_1	2.57	5.00
	S_2	11.43	13.00
	S_3	18.86	21.14
	S_4	3.57	7.29

Table 7

Observee Reactivity: Valence (2) x Teachers Within Valence (2) x Intervals (2) x Days (7) Repeated Measures Analysis of Variance For Appropriate Student Verbalizations

Source	df		MS		F	
V	1		159 47		2.54	(1.37)[a]
T(V)	2	(14)	62.79	(115.99)	0.50	
S(T(V))	12		124.87			
I	1		51.11		0.13	
VI	1		267.97		0.70	
IT(V)	2		380.79		3.19	
IS(T(V))	12		119.44			
D	6		26.80		1.08	(1.08)[b]
VD	6		31.05		1.25	(1.25)[b]
DT(V)	12	(84)	24.85	(24.92)	1.00	
DS(T(V))	72		24.93			
ID	6		12.49		0.56	(.33)[c]
VID	6		30.05		1.36	(.79)[c]
IDT(V)	12	(84)	22.12	(37.92)	0.55	
IDS(T(V))	72		40.55			

Note. Parentheses indicate calculations using pooled error term.

[a]Test of V with pooled error term (T(V) and S(T(V)).

[b]Test of D and VD with pooled error term (DT(V) and DS(T(V)).

[c]Test of ID and VID with pooled error term (IDT(V) and IDS(T(V)).

Table 8

Frequency of Appropriate and Inappropriate Student Verbalizations Made By the Students in Each Teacher's Classroom During Baseline (I_1) and Teacher Observation (I_2)

	Appropriate Verbalizations		Inappropriate Verbalizations	
	Baseline (I_1)	Teacher Obs. (I_2)	Baseline (I_1)	Teacher Obs. (I_2)
Teacher 1	6.07	12.86	3.86	4.21
Teacher 2	7.79	7.26	8.18	8.61
Teacher 3	6.00	8.50	1.29	1.96
Teacher 4	8.86	3.89	8.93	11.61

Table 9

Observee Reactivity: Valence (2) x Teachers Within Valence (2) x Intervals (2) x Days (7) Repeated Measures Analysis of Variance For Inappropriate Student Verbalizations

Source	df		MS		F	
V	1		4.02		0.00	
T(V)	2		1311.57		6.55**	
S(T(V))	12		200.29			
I	1		60.07		4.29	(1.23)[a]
VI	1		23.14		1.65	(0.47)[a]
IT(V)	2		14.02		0.26	
IS(T(V))	12	(14)	54.72	(48.90)		
D	6		77.66		4.60*	(2.56*)[b]
VD	6		39.04		2.31	(1.29)[b]
DT(V)	12		16.89		0.52	
DS(T(V))	72	(84)	32.51	(30.28)		
ID	6		14.42		0.61	(0.47)[c]
VID	6		30.73		1.29	(1.00)[c]
IDT(V)	12		23.82		0.74	
IDS(T(V))	72	(84)	31.99	(30.82)		

Note. Parentheses indicate calculations using pooled error term.

[a]Test of I and VI with pooled error term (IT(V) and IS(T(V)).

[b]Test of D and VD with pooled error term (DT(V) and DS(T(V)).

[c]Test of ID and VID with pooled error term (IDT(V) and IDS(T(V)).

*$p < .05$.

**$p < .01$.

Table 10

Observer Reactivity: Valence (2) x Teachers Within Valence (2) x Intervals (2) x Days (7) Repeated Measures Analysis of Variance For Positive Teacher Behavior

Source	df		MS		F	
V	1		0.36		0.00	
T(V)	2		140.31		11.32**	
S(T(V))	12		12.39			
I	1		214.11		4.09	
VI	1		0.36		0.01	
IT(V)	2		52.29		5.28*	
IS(T(V))	12		9.89			
D	6		6.09		1.28	(1.12)[a]
VD	6		6.60		1.38	(1.21)[a]
DT(V)	12		4.77		0.86	
DS(T(V))	72	(84)	5.56	(5.45)		
ID	6		5.56		1.50	(0.95)[b]
VID	6		1.04		0.28	(0.18)[b]
IDT(V)	12		3.71		0.60	
IDS(T(V))	72	(84)	6.22	(5.86)		

Note. Parentheses indicate calculations using pooled error term.

[a]Test of D and DV with pooled error term (DT(V) and DS(T(V)).

[b]Test of ID and VID with pooled error term (IDT(V) and IDS(T(V)).

*$p < .05$.

**$p < .01$.

Table 11

Mean Frequency Across Students of Positive, Negative, and Instruction Behaviors Exhibited By Each Teacher During Baseline (I_1) and Teacher Observations (I_2)

	Positive Behaviors		Negative Behaviors		Instruction Behaviors	
	Baseline (I_1)	Teacher Obs. (I_2)	Baseline (I_1)	Teacher Obs. (I_2)	Baseline (I_1)	Teacher Obs. (I_2)
Teacher 1	2.82	6.43	0.82	1.11	6.64	14.18
Teacher 2	2.57	3.03	0.39	0.57	6.50	5.82
Teacher 3	3.43	6.43	0.60	0.39	6.04	8.68
Teacher 4	1.96	2.71	0.54	0.71	3.14	4.04

Table 12

Observer Reactivity: Valence (2) x Teachers Within Valence (2) x Intervals (2) x Days (7) Repeated Measures Analysis of Variance For Negative Teacher Behavior

Source	df		MS		F	
V	1		1.45		0.42	(0.62)[a]
T(V)	2	(14)	3.47	(2.33)	1.63	
S(T(V))	12		2.09			
I	1		0.64		1.11	(0.76)[b]
VI	1		0.88		1.51	(1.03)[b]
IT(V)	2	(14)	0.58	(0.85)	0.65	
IS(T(V))	12		0.90			
D	6		0.26		0.25	(0.39)[c]
VD	6		1.41		1.37	(2.13)[c]
DT(V)	12	(84)	1.03	(0.66)	1.71	
DS(T(V))	72		0.60			
ID	6		0.94		1.11	(1.14)[d]
VID	6		0.40		0.47	(0.48)[d]
IDT(V)	12	(84)	0.84	(0.82)	1.03	
IDS(T(V))	72		0.81			

Note. Parentheses indicate calculations using pooled error term.

[a]Test of V with pooled error term (T(V) and S(T(V)).

[b]Test of I and VI with pooled error term (IT(V) and IS(T(V)).

[c]Test of D and VD with pooled error term (DT(V) and DS(T(V)).

[d]Test of ID and VID with pooled error term (IDT(V) and IDS(T(V)).

Table 13

Observer Reactivity: Valence (2) x Teachers Within Valence (2) x Intervals (2) x Days (7) Repeated Measures Analysis of Variance For Instruction Teacher Behavior

Source	df		MS		F	
V	1		442.97		0.98	
T(V)	2		451.60		12.90**	
S(T(V))	12		35.01			
I	1		378.04		1.53	
VI	1		38.61		0.16	
IT(V)	2		246.89		9.03*	
IS(T(V))	12		27.34			
D	6		4.31		0.20	(0.30)[a]
VD	6		22.75		1.06	(1.61)
DT(V)	12		21.54		1.67	
DS(T(V))	72	(84)	12.89	(14.12)		
ID	6		14.82		1.23	(0.90)[b]
VID	6		8.79		0.73	(0.53)[b]
IDT(V)	12		12.07		0.70	
IDS(T(V))	72	(84)	17.16	(16.43)		

Note. Parentheses indicate calculations using pooled error term.

[a]Test of D and VD with pooled error term (DI(V) and DS(T(V)).

[b]Test of ID and VID with pooled error term (IDT(V) and IDS(T(V)).

*$p < .05$.

**$p < .01$.

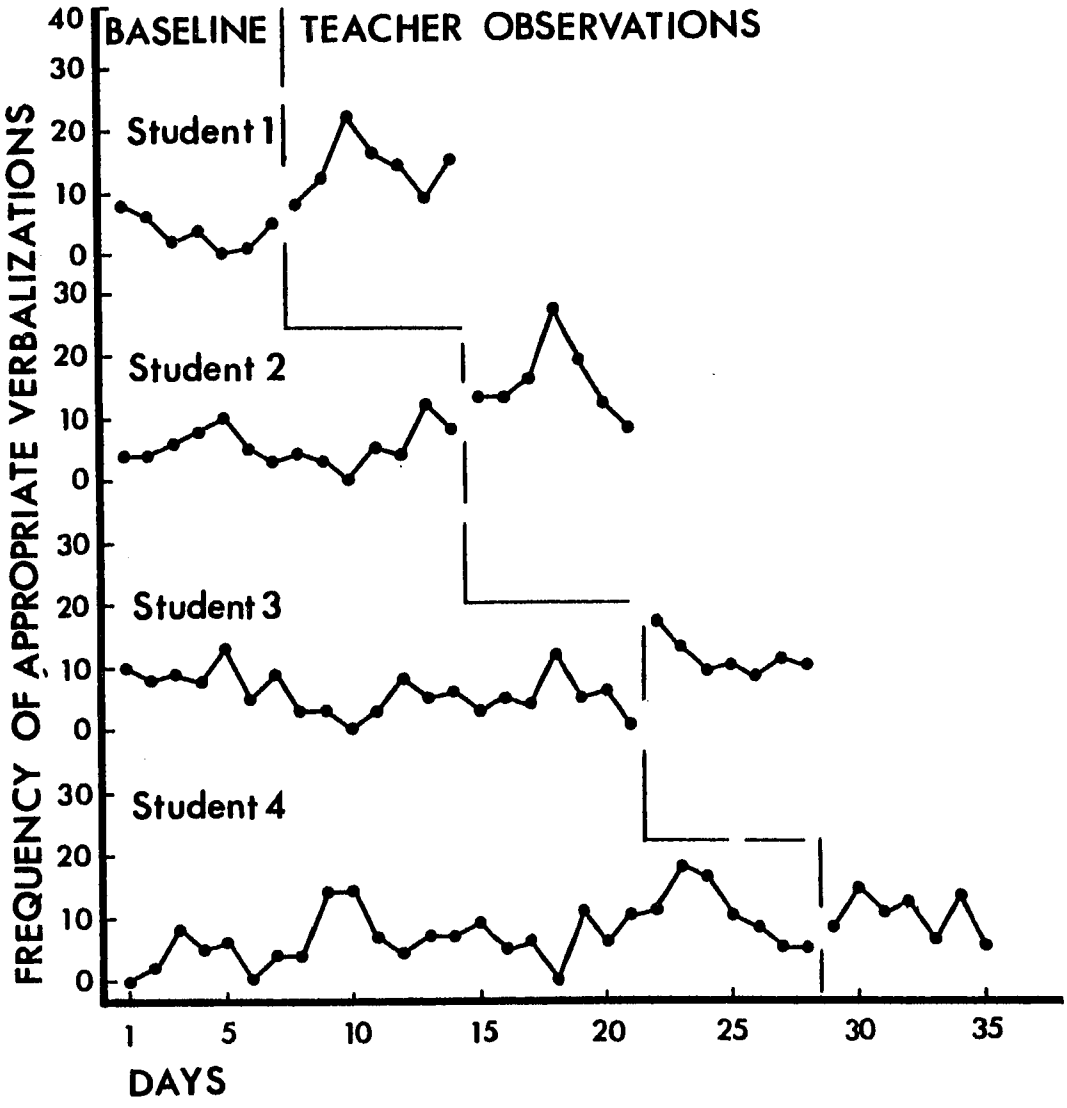

Figure 1a. Frequency of appropriate student verbalizations exhibited by each student in Teacher 1's classroom during baseline and teacher observations.

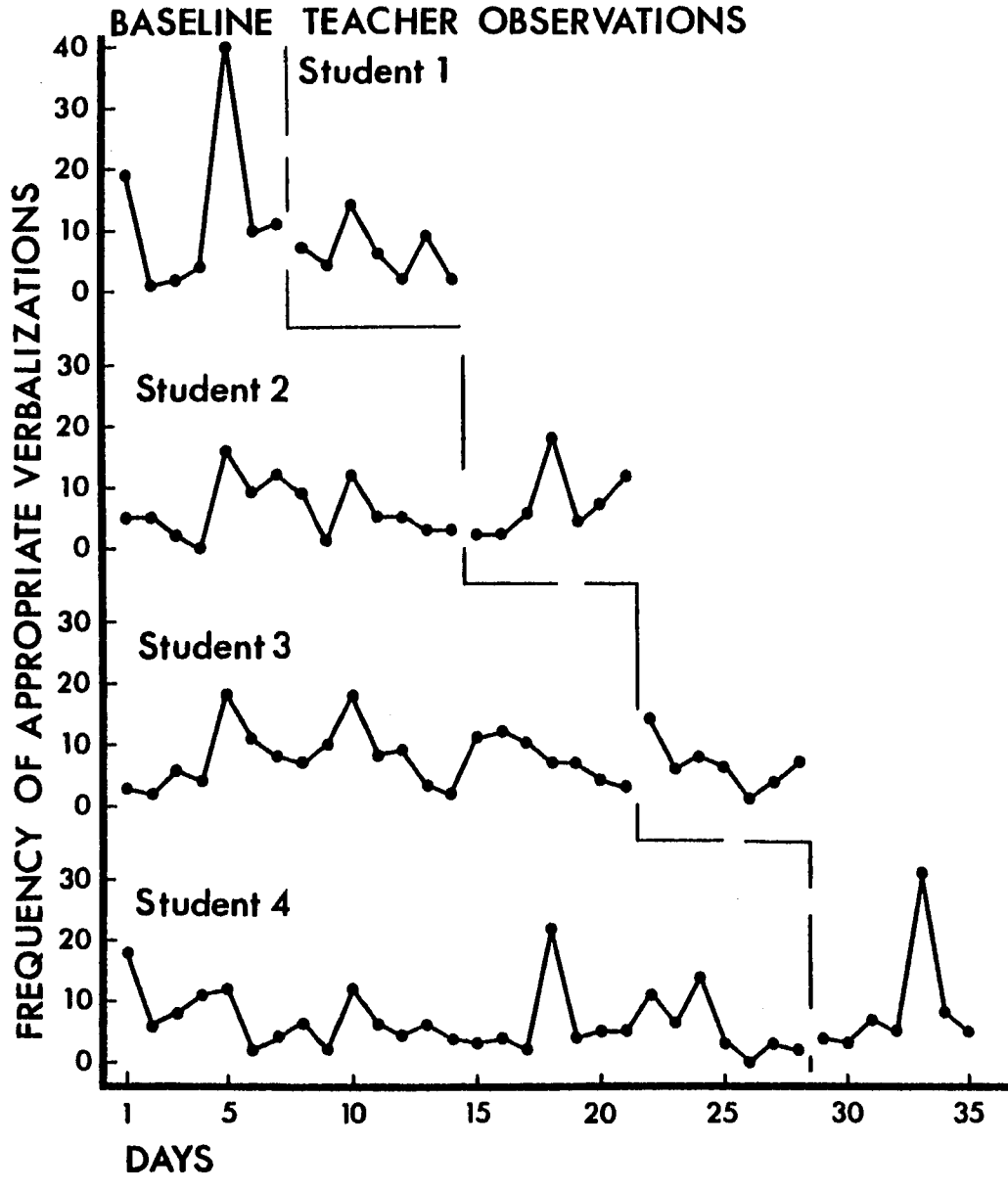

Figure 1b. Frequency of appropriate student verbalizations exhibited by each student in Teacher 2's classroom during baseline and teacher observations.

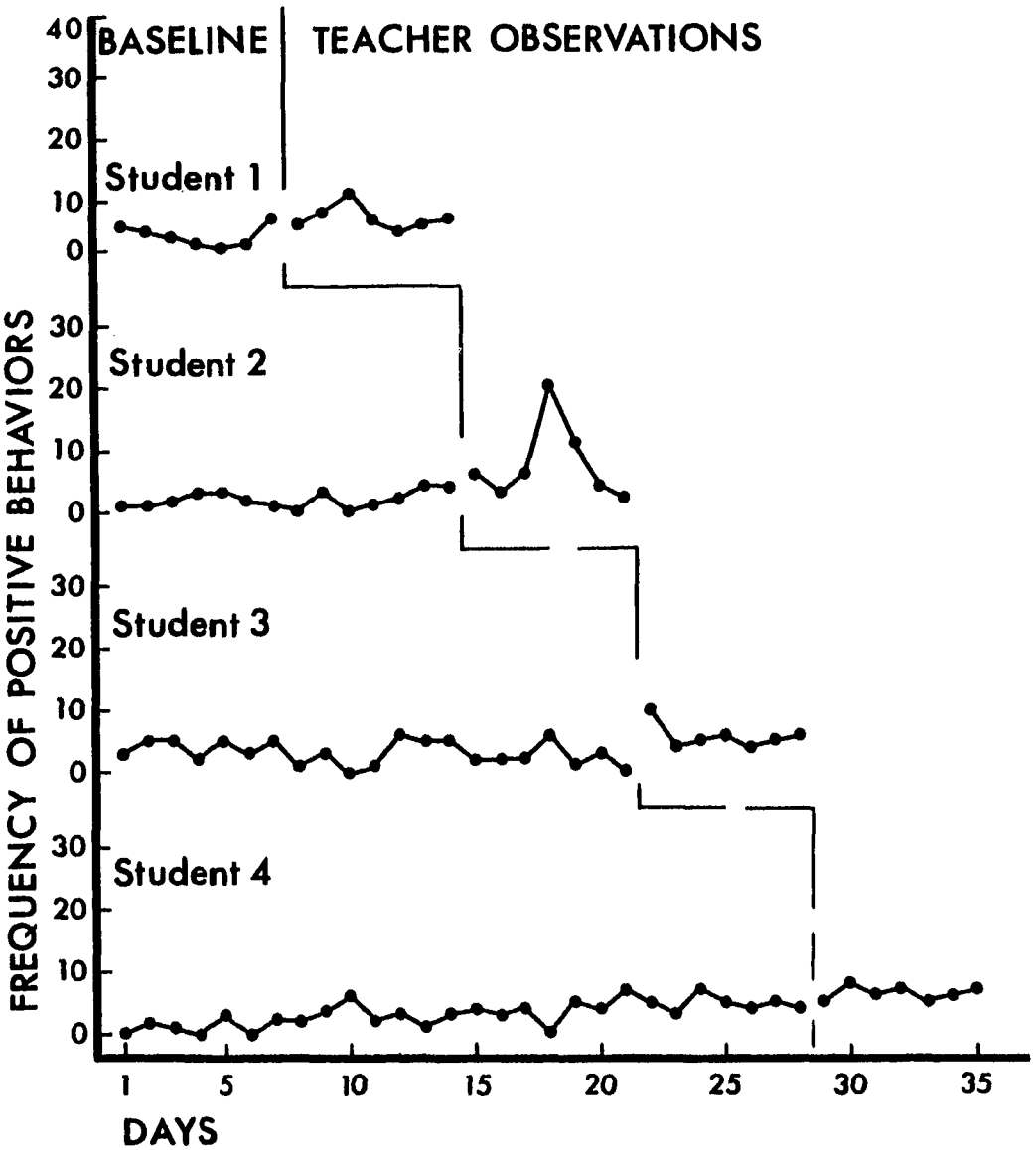

Figure 2a. Frequency of positive behaviors exhibited daily by Teacher 1 to each student during baseline and teacher observations.

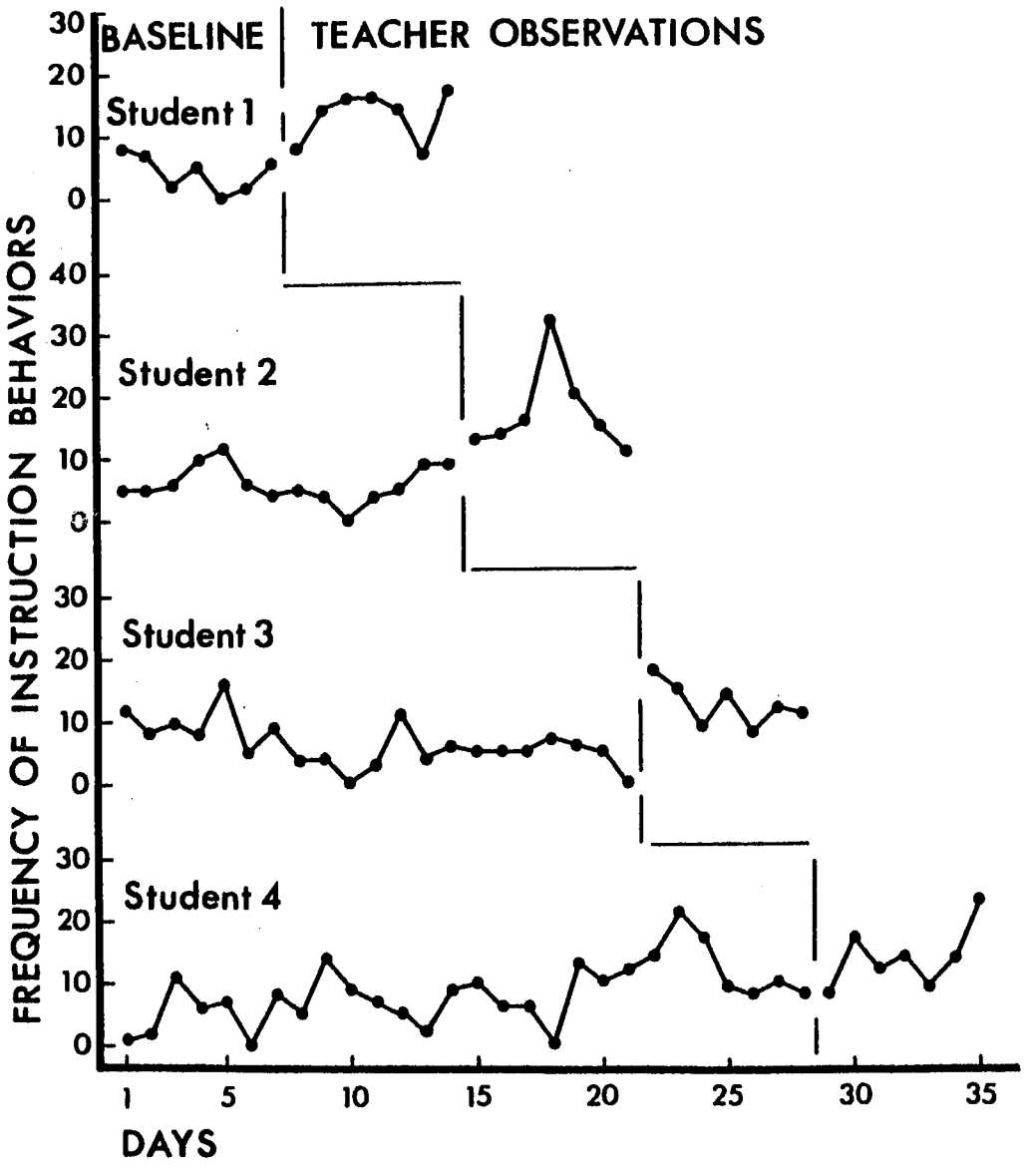

Figure 2b. Frequency of instruction behaviors exhibited daily by Teacher 1 to each student during baseline and teacher observations.

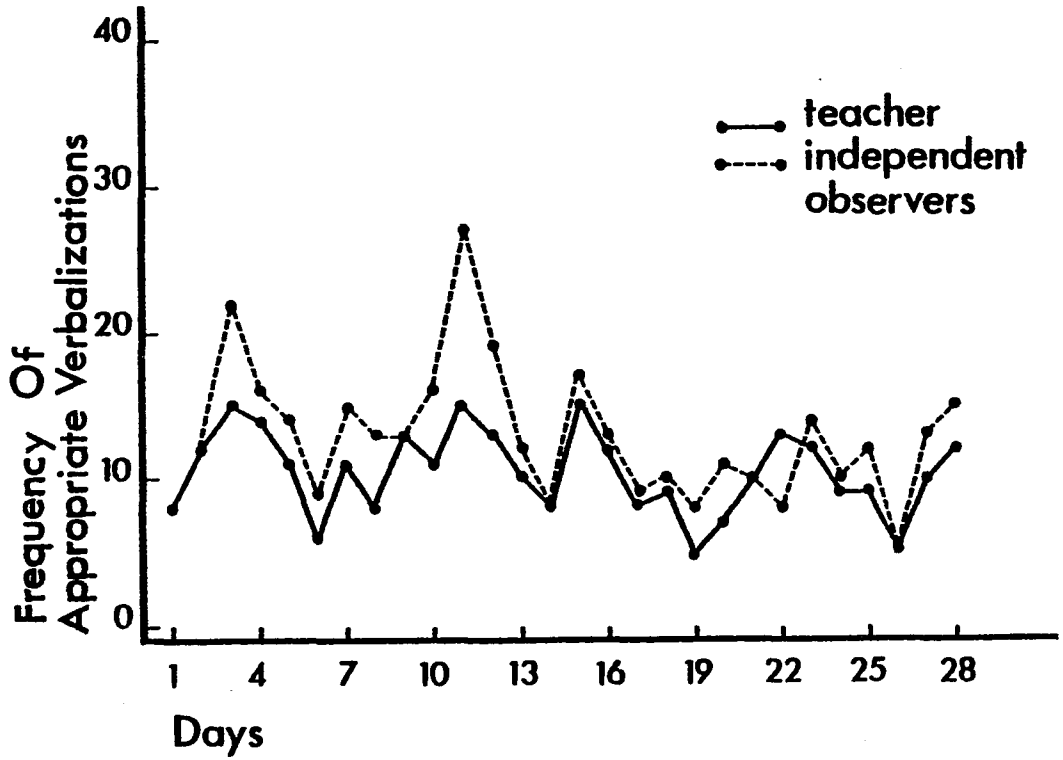

Figure 3a. Daily frequency of appropriate student verbalizations as recorded by Teacher 1 and the independent observers.

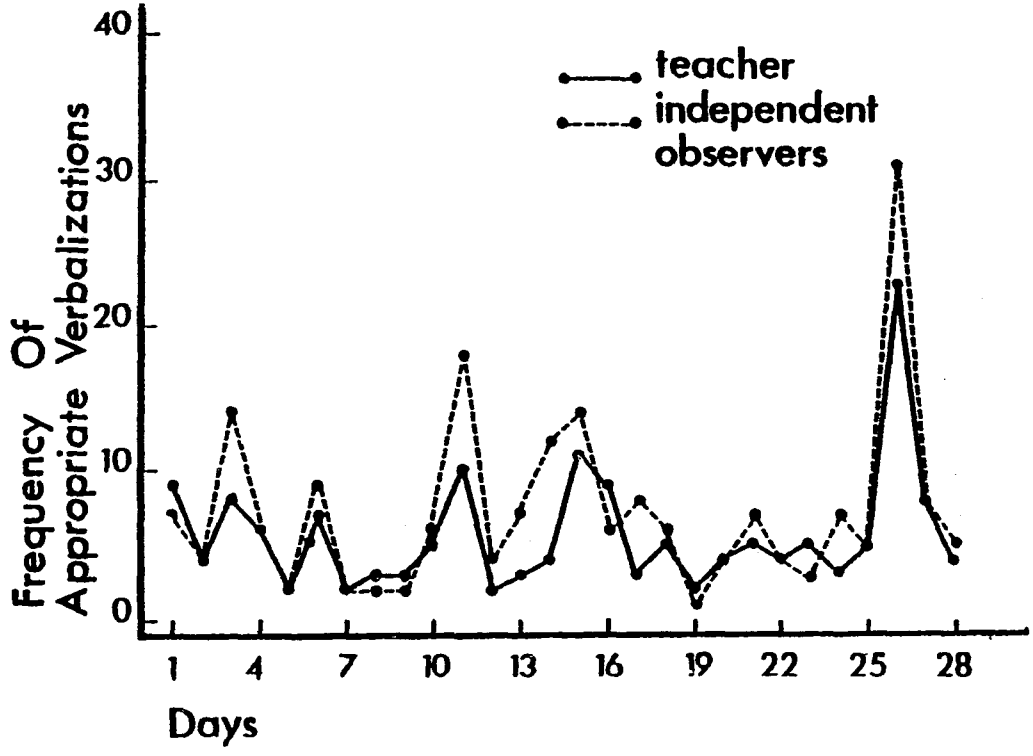

Figure 3b. Daily frequency of appropriate student verbalizations as recorded by Teacher 2 and the independent observers.

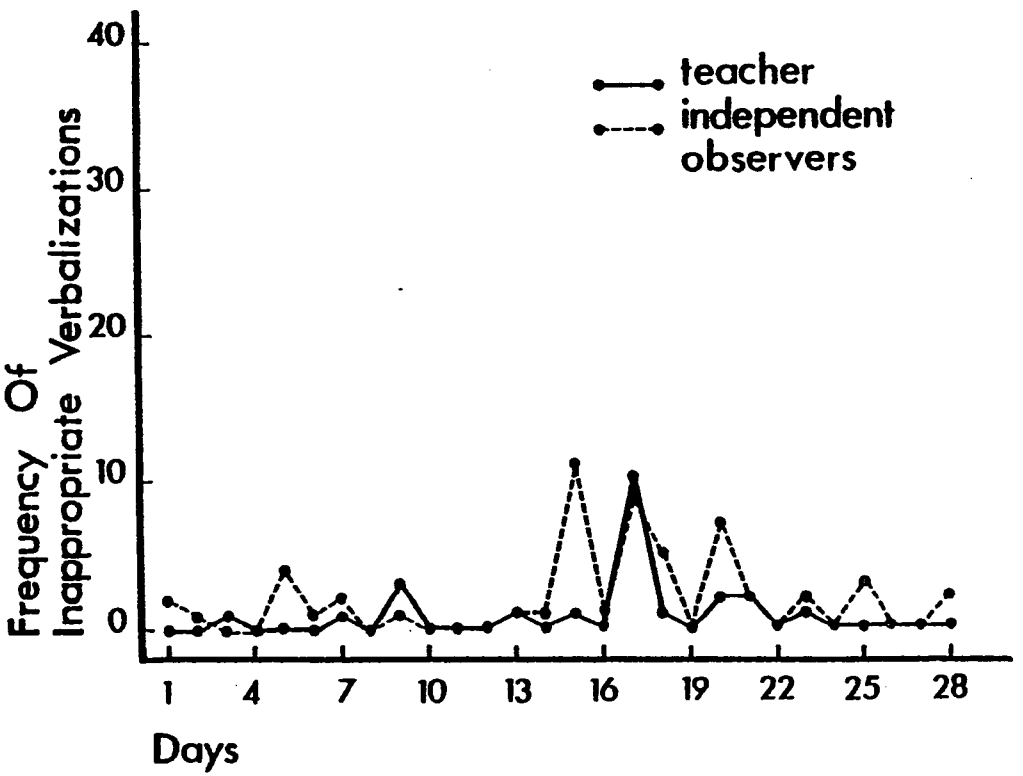

Figure 3c. Daily frequency of inappropriate student verbalizations as recorded by Teacher 3 and the independent observers.

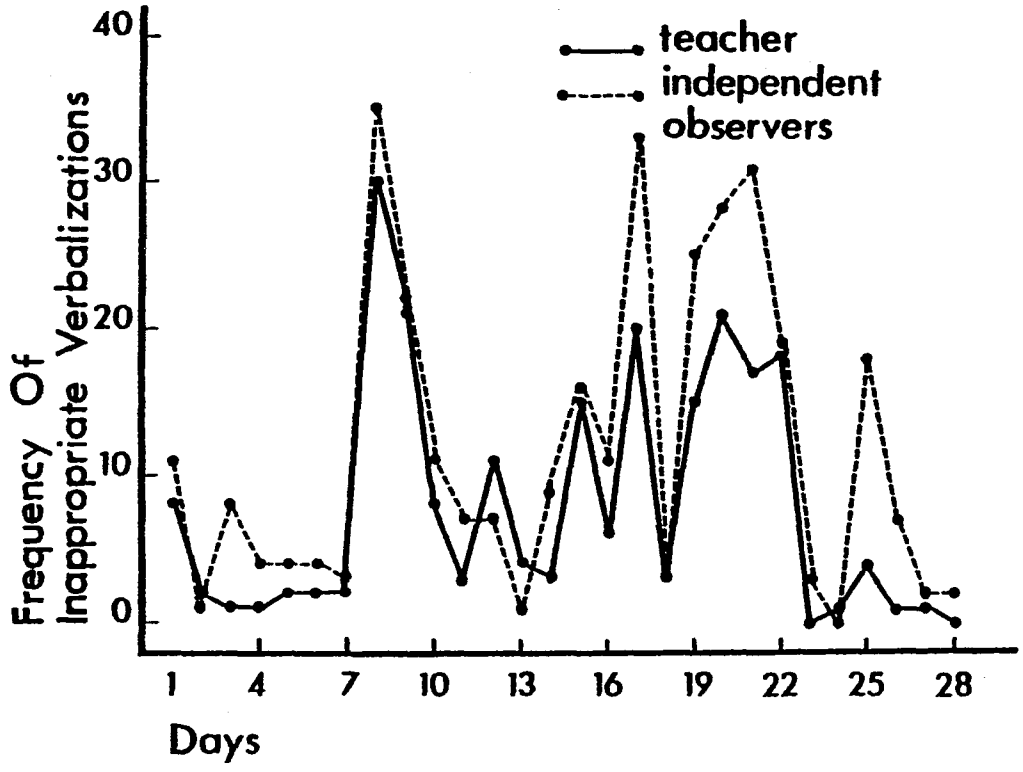

Figure 3d. Daily frequency of inappropriate student verbalizations as recorded by Teacher 4 and the independent observers.

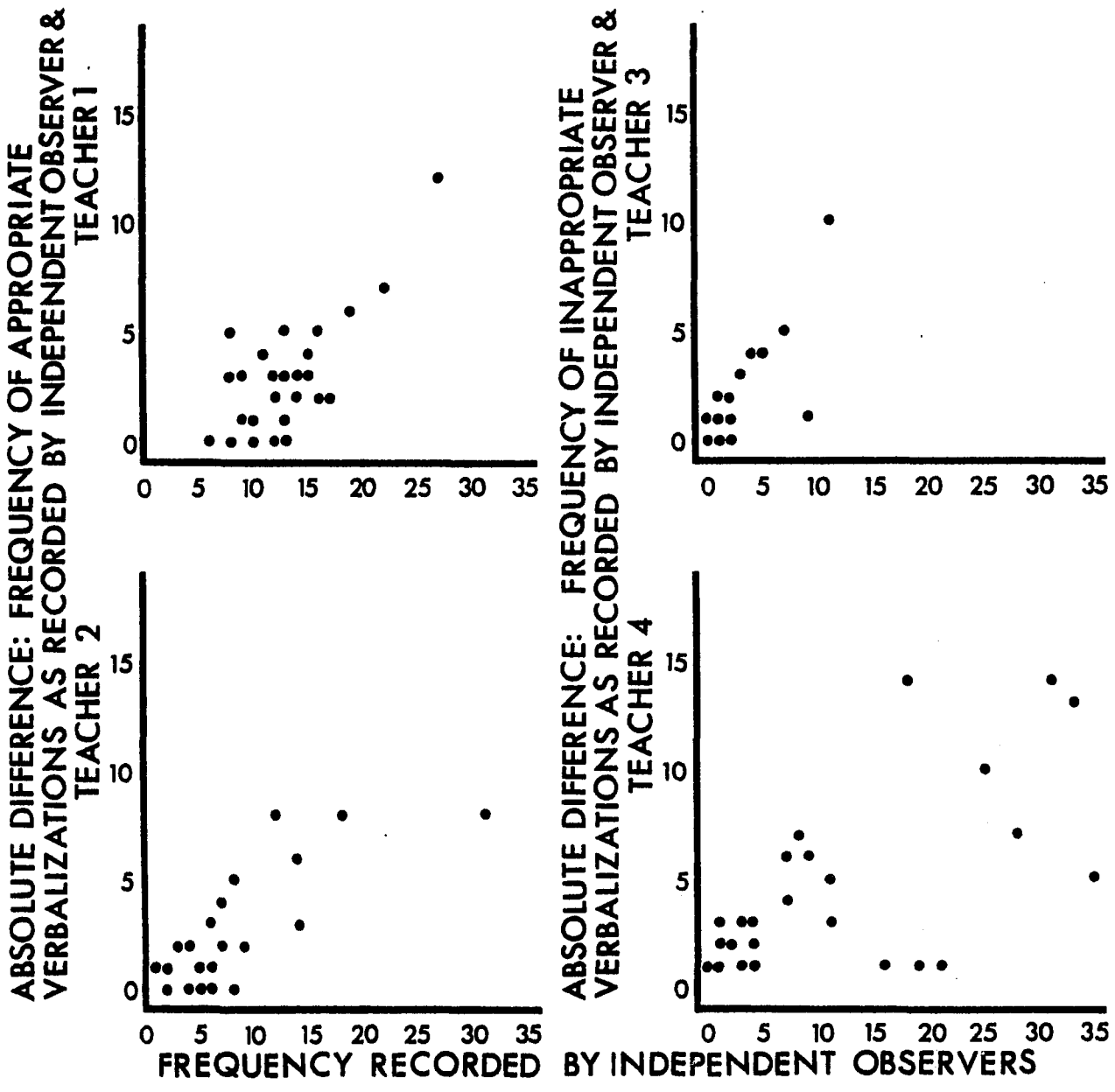

Figure 4. Relationship between response frequency as recorded by the independent observers and teacher reliability (absolute difference between response frequencies as recorded by each teacher and the independent observers).