75-3215

VASTBINDER, Sherry Stienecker, 1938-
THE EFFECTS OF KNOWLEDGE OF PERFORMANCE ON THE BEHAVIOR OF EDUCABLE MENTALLY RETARDED STUDENTS AND THEIR TEACHERS.

The Ohio State University, Ph.D., 1974
Education, special

Xerox University Microfilms, Ann Arbor, Michigan 48106

THE EFFECTS OF KNOWLEDGE OF PERFORMANCE ON THE BEHAVIOR OF EDUCABLE MENTALLY RETARDED STUDENTS AND THEIR TEACHERS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Sherry S. Vastbinder, B.Sc., M.A.

* * * * *

The Ohio State University

Reading Committee:

Thomas M. Stephens, Ed. D.
John O. Cooper, Ph. D.
Claribel Taylor, Ph. D.
A. Lee Parks, Ph. D.

Approved By

T. M. Stephens

Adviser
Faculty for Exceptional Children

ACKNOWLEDGMENTS

The author gratefully acknowledges the contributions of her doctoral committee in the completion of this manuscript. Special appreciation goes to Dr. Thomas M. Stephens, major adviser, for his personal interest and assistance and especially for the quality of the doctoral program which he maintains; a debt of gratitude is owed Dr. John O. Cooper for his painstaking efforts in editing and revision and for the positive reinforcement he liberally supplied; personal thanks to Dr. Claribel Taylor, who was both a teacher and a friend; and a special thank you to Dr. Lee Parks for his constructive advice.

Much gratitude is felt for the personal sacrifice of my family: my husband, Walter, whose steadfast encouragement and support made this accomplishment possible; and our children, Joyce, Marvin, Erin and Eric, who assumed the many responsibilities which I neglected.

A special thank you is extended to both the student and teacher-subjects who so willingly participated in this study.

Finally, I acknowledge my indebtedness to friends and family who supported me with their prayers, and to my God, Who holds me in the hollow of His hand.

VITA

July 5, 1938 . . .	Born - Lima, Ohio
1967-68	Special Education Teacher, Trainable Mentally Retarded, Springfield, Ohio
1968-69	Special Education Teacher, Educable Mentally Retarded, Springfield, Ohio
1969-70	Special Education Teacher, Educable Mentally Retarded, Fairborn, Ohio
1970	B.Sc., Psychology, The Ohio State University, Columbus, Ohio
1971	M.A., School Psychology, The Ohio State University, Columbus, Ohio
1971-72	School Psychologist Intern, Montgomery County, Ohio
1971-72	Instructor--Evening Sessions, Wittenberg University, Springfield, Ohio
1972-74	Graduate Research Associate, The National Center on Educational Media and Materials for the Handicapped, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Exceptional Children

Studies in Learning and Behavior Disorders. Professors Thomas M. Stephens and John O. Cooper.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION TO THE STUDY

Accountability has become a major subject of educational concern within the last ten years.[1] One who reads the literature (e. g. Tyler, 1971, pp. 1-6; Vergason, 1973) soon finds that scarcely any other topic in the history of education has been as widely acknowledged, as emotionally charged, or as universally accepted by parents, taxpayers, and legislators. The accountability demands made by Congress in the Elementary and Secondary Education Act (1965); the legal action initiated by parents relative to their children's placement in classes for the educable mentally retarded; and the actions taken by various state legislatures to insure educational assessment, reflect the growing concern with the evaluation of our education process (Milliken, 1971, pp. 18-20; Peterson,1971, pp. 20-24). Most individuals addressing the accountability issue (Lessinger, 1971, pp. 7-15; Nyquist, 1971, pp. 24-27; Tyler, 1971, pp. 1-6; Porter, 1971, pp. 42-51; Reddick, 1972; Vergason, 1973) suggest that responsibility of the accountability

---

[1]For the purposes of this study, accountability is defined as the critical analysis of the teaching-learning process.

process rests with the teacher. Many further suggest that reports on school achievement should be available to the community at public meetings. In fact, Lessinger (1971, pp. 7-15) writes that communities can insist that teachers become accountable for relating the teaching process and teaching procedure to results as expressed by changes in pupil behavior.

Rationale

The effectiveness of teachers is not a new concern. A large number of studies of teacher effectiveness have been conducted. Barr (1948), for example, summarized some 150 studies completed up to 1948. In addition, Domas and Tiedman (1950) reported extensive bibliographies of related studies. Inspection of these attempts indicate that, despite all efforts expended, very little real progress was made.

The basic pattern for most studies was the correlation of measures of various teacher characteristics with either ratings of teacher competency or with average measures of the limited areas of pupil growth. Orleans, Clarke, Ostreicher and Standlee (1952) concluded that the types of approaches to the problem were not fruitful ones in the sense that they did not furnish adequate answers to questions about teacher competencies. Many of the studies summarized by Barr (1948) made use of proximate criteria in the form of ratings of teachers as though they were the ultimate criteria. Several authors

(Orleans, Clarke, Ostreicher and Standlee, 1952; Morsh, Burgess and Smith, 1953; Mitzel and Medley, 1954; Medley and Mitzel, 1959) pro-posed that the ultimate criteria for the effectiveness of the teacher's performance are the changes which take place in the behavior of the pupils. Unfortunately, research dealing with the measurement of teacher effectiveness and measures of the degree to which pupils are learning from the teacher have little in common (Hellfritsch, 1945; La Duke, 1945; Jayne, 1945; Gotham, 1945; Brookover, 1945; Lins, 1946 and Anderson, 1954). The task then becomes one of devising techniques which will measure changes in pupil behavior.

In 1968, B. F. Skinner (The Technology of Teaching) attended to the task of increasing teacher effectiveness based on three assumptions: (a) student failures are caused by teaching failures; (b) teacher competency is measured in terms of demonstrated student accomplishment; and (c) stating objectives in behavioral terms promotes a clearer communication of intended outcomes. Knowledge of intended outcomes, in turn, provides the criteria for evaluation. Skinner (1968) was suggesting that teachers should be responsible for measuring their teaching effectiveness in terms of pupil progress. He further suggested that this could be done by comparing behavioral objectives with actual pupil performance.

During the same era, emphasis in educational research began shifting from the investigation of general questions to the investigation of specific procedures such as observation and recording procedures (Hall and Broden, 1967; Thomas, Becker and Armstrong, 1968; Hall, Lund and Jackson, 1968); time-sampling procedures (MacKenzie, Clark, Wolf, Kothera and Benson, 1968); applied behavior analysis designs (Baer, Wolf and Risley, 1968); new instructional models (Peter, 1965; Hewett, 1967; Hall, 1972) together with their effects on classroom groups and individual children (Hall, 1972, pp. 403-414). The introduction of two new basic research designs in education, "reversal" and "multiple baseline" scientific verification procedures, (Baer, Wolf and Risley, 1968; Hall, 1972) furthered the study of the effectiveness of individual teachers in individual classrooms. Both designs are equipped to handle single subjects or single groups. In addition, both designs allow teachers to make probability statements that their instructional procedures are responsible for changes in student behavior.

A further consideration related to the accountability issue is the effect that evaluation of the teaching-learning process will have on the teacher, the students, and the community of which they are a part. Recording and graphing behavior is basic to self-accountability. By graphing data, the teacher creates a visual representation of behavior which is easily interpreted by students and parents. Not only is this an effective evaluation technique (Hall, 1972, pp. 403-414), but a review of research indicates that knowledge of performance has some interesting effects on the subjects involved.

Studies related to knowledge of performance suggest that a variety of variables may confound experiments wherein knowledge of performance is being manipulated. However, analysis of student behavior and knowledge of results suggests that knowledge may be a reinforcer when uncertainty or probability of emitting an incorrect response is high, or where confidence is low (Geis and Chapman, 1972).[2]

Given the last suggestion, special education teachers might find the use of knowledge of results with handicapped learners a worthwhile technique because the students' experience with learning may have been frustrating and their self-confidence low in learning situations. These teachers may also find useful the continually gathered information as to how, when, and why knowledge of one's own performance in a learning situation becomes reinforcing, and contributes to more effective learning.

Whether sharing knowledge of performance with students does or does not improve performance, sharing knowledge of student performance with the public still remains a conceivable method of assuring the public that schools are evaluating effectiveness of their teaching-learning process. The assurance of this evaluation is basic to providing accountability in education.

---

[2]In this study, knowledge of performance and knowledge of results are used synonymously.

Statement of the Problem

The major purpose of this experiment was to study the effects of knowledge of arithmetic performance of educable mentally retarded students on both the teachers' and students' behavior. The following research questions were proposed by the investigator:

(1) Will knowledge of student performance significantly change student and/or teacher behavior?

(2) Will the public display of student performance significantly change student and/or teacher behavior?

Null Hypotheses

Out of the above stated research questions, the following null hypotheses were tested for tenability.

(1) There will be no statistically significant change ($\alpha = .05$) in the number of correct arithmetic responses emitted by three low achieving students when knowledge of daily quiz scores are recorded and graphed and shared with the student by the teacher.

$$Ho_1: \overline{X}_{A_1} = \overline{X}_{A_2}$$

(2) There will be no statistically significant change ($\alpha = .05$) in the number of correct arithmetic responses emitted by three low achieving students when the recorded and graphed results of number of correct responses on daily arithmetic quizzes are publicly displayed on the classroom bulletin board.

$$Ho_2: \bar{X}_{A_1} = \bar{X}_{A_2} = \bar{X}_{A_3}$$

(3) There will be no statistically significant change ($\alpha = .05$) in the amount of time the teacher spends in arithmetic lessons when she/he records, graphs, and shares the number of correct responses on daily arithmetic quizzes with three low achieving students.

$$Ho_3: \bar{X}_{B_1} = \bar{X}_{B_2}$$

(4) There will be no statistically significant change ($\alpha = .05$) in the amount of time the teacher spends in arithmetic lessons when the number of correct responses on daily arithmetic quizzes of three low achieving students are recorded, graphed and publicly displayed on the classroom bulletin board.

$$Ho_4: \bar{X}_{E_1} = \bar{X}_{B_2} = \bar{X}_{B_3}$$

<u>Definitions of Means</u>

$\bar{X}_{A_1}$: the mean of the number of correct responses emitted during the baseline phase of the study.

$\bar{X}_{A_2}$: the mean of the number of correct responses emitted during the phase wherein daily quiz scores are recorded, graphed and shared with the student.

$\bar{X}_{A_3}$: the mean of the number of correct responses emitted during the phase wherein daily quiz scores are recorded, graphed, shared with the student and publicly displayed on the classroom bulletin board.

$\overline{X}_{B_1}$: the mean of the amount of time spent in arithmetic lessons during baseline phase of the study.

$\overline{X}_{B_2}$: the mean of the amount of time spent in arithmetic lessons during the phase wherein daily quiz scores are recorded, graphed and shared with the student.

$\overline{X}_{B_3}$: the mean of the amount of time spent in arithmetic lessons during the phase wherein daily quiz scores are recorded, graphed, shared with the student and publicly displayed on the classroom bulletin board.

Limitation/Delimitation

(1) The population was limited to educable mentally retarded students and their teachers. This was done because results of studies indicate that knowledge of results may be a reinforcer when uncertainty or probability of emitting an incorrect response is high, or where confidence is low (Geis and Chapman, 1972). The inclusion of better students would probably affect the findings of this study as knowledge of results is apparently ignored by students who perform academically with relative certainty about their performance (Lumsdaine, 1961; Holland, 1965). A second consideration in selecting educable mentally retarded students is the call for

accountability in education, particularly in special classes for the mildly retarded (Dunn, 1968).

Delimitation: To further limit the above listed limitations, educable mentally retarded students were selected from one particular community due to constraints of time, money, and availability. Also, the teachers of these students were willing to volunteer as participants in this study.

(2) The population was limited to teachers of the educable mentally retarded. The behaviors of the teachers were also studied because results of research indicate knowledge of student performance changes teacher behavior (Bapst, 1972; Novak and Moser, 1972).

Delimitation: To further limit the above listed limitation, the teachers were selected from the same school district as the educable mentally retarded students because their teaching behavior was recorded simultaneously with their students performance.

(3) Sample size was limited by the number of special education units within the chosen school district. N = 12 (teachers)

Delimitation: To further limit the above listed limitation, the use of the multiple baseline design with a replication group and a control group further limited the number of teacher participants to three. Due to this design, the number of student participants was further limited to nine.

(4) Subjects were elementary and junior high school students between the ages of seven years-four months and fourteen years-seven months. Both individualizing instruction and taping arithmetic sessions were necessary in this study. Since senior high school students were not instructed as a group due to work schedules, it was inconvenient to measure teacher behavior by taping arithmetic sessions. Therefore, the senior high educable mentally retarded students were not included in the sample.

Operational Definitions

In order to avoid any undue misunderstanding, the investigator has chosen to operationally define the following terms.

Low Achiever: A student who is achieving two to three years or more below the grade level at which the child would be placed based on his chronological age.

Feedback: Feedback has been defined in this study as knowledge of the number of correct arithmetic responses on daily quizzes as recorded on a graph.

Stable Behavior Rate: Stable behavior rate has been defined in this study as three or more successive sessions in which the number of correct arithmetic responses do not fluctuate more than two points from the mean of these responses.

CHAPTER II

REVIEW OF RELATED LITERATURE

Accountability procedures in relation to behavior analysis are the main focus of this chapter. The following topics are presented: personal accountability for teachers; formal evaluation techniques; informal evaluation techniques; and a detailed review of the literature relevant to the effects of knowledge of performance.

Personal Accountability for Teachers

Porter (1971, pp. 42-51) defines accountability as the guarantee that all students, without respect to race, income, or social class, will acquire the minimum school skills necessary to take advantage of the choices that accrue upon successful completion of public schooling. Lessinger (1971, pp. 7-15) expands the definition by including a regular public report by independent reviewers along with demonstrated student accomplishment promised for the expenditure of resources. Self or personal-accountability as practiced by teachers then would consist of measuring their own performance, demonstrating student accomplishment, and making public reports of these data.

Bain, (1972) immediate past president of the National Education Association, said that self-governance must precede total accountability to the public by the profession. Reddick (1972) writes that self-accountability, by a dedicated and conscientious professional, is the most valuable type of accountability for which the public could ever hope.

What are some steps then which teachers should take to assure personal accountability? The teacher may use two kinds of student evaluation techniques. These are generally referred to as "formal" and "informal" evaluation techniques. The three basic differences in these techniques are: (1) the manner in which the two are constructed, (2) the meaning of results obtained from each, and (3) the uses for which the results are appropriate (Stephens, 1974).

Formal Evaluation Techniques

For the most part, formal evaluation procedures have been used to measure pupil achievement. Formal procedures include such measures as normed (standardized) tests and inventories, six weeks' grades (A, B, C, D, F) and final examinations. These tests compare the student's performances with other students and assume that pupil achievement is normally distributed. Normed tests are useful for curriculum development, for counseling students on future planning, and for evaluating groups of students. Group tests evaluating group instructional practices, (i.e., everyone receiving the same instruction in the same amount of time) reveal differences between pupils and rank them from high to low. Only a proportionate number reach a mastery level.

Grade scores, commonly obtained from standardized achievement tests, are useful when comparing a student's position to a reference group, but are too general and, consequently, have limited instructional value for teachers (Stephens, 1970, p. 64). Relationships have been observed between a pupil's score on a normed test and his classroom performance. However, normed evaluation is at best indirectly related to the pupil's daily assignment. When measurement is not directly related to instruction, improvement in a pupil's day-to-day performance could go undetected. Unless academic and social behaviors are explicitly defined and measured, no one will know whether or not that skill has been mastered, or with what success (Lovitt, Schaff and Sayre, 1972, pp. 416-432).

The most fruitful use of evaluation information is not only to determine students' achievement, but to indicate instructional changes that might be warranted. It is essential to gather instructionally relevant information by comparing student performance with previously defined instructional objectives. This is accomplished by using informal evaluation techniques.

Informal Evaluation Techniques

As stated earlier, greater emphasis is generally placed on the use of formal evaluation. But informal evaluation of instructional objectives, that is (1) criterion-referenced measures, (2) observational recordings, and (3) direct measurement of permanent products, should also be included insofar as it can be readily applied by the classroom teacher.

Student progress should be systematically and frequently evaluated with criterion-referenced measures. The criterion-referenced measure is a fixed standard which allows the teacher to compare the individual's performance to a pre-determined criterion rather than to other students' performance. As a general rule, any content that has identifiable skills and concepts that can be translated into responses is directly assessable by using criterion measures. These measures represent an instructional tool that should be part of the teaching process. Its use facilitates individualized instruction and enables teaching to become more accountable (Stephens, 1974).

Behavior observation is important to the total evaluation process. When a human observer looks at behavior and makes a record of what he sees as it occurs, he is engaging in observational recording. There are several types of observational recording (Hall, 1971, pp. 1-3): (1) continuous recording--sometimes called an anecdotal record; (2) event recording--a frequency count of events as they occur; (3) duration recording--recording the elapsed time of a specific behavior; (4) interval recording--recording the occurrence of behavior during a number of equal time periods; and (5) time sampling--recording behavior only at the end of each specified time interval. Brown (1972, p. 388) stipulates that behavior observations, as a portion of informal evaluation techniques, can and should be used continuously if the instructional process is to be judged as successful.

The third informal evaluation technique is the direct measurement of permanent products. Examples of permanent products include written arithmetic computation, written spelling words, written alphabet letters, colorings, building puzzles, stringing beads, and stacking blocks. These products are tangible and can be measured subsequent to the student's behavior (Cooper, 1974). This procedure is concrete and direct and is one way for teachers, parents, and children to measure student performance (Brown, 1972, p. 392). A permanent product can be observed and counted and lends itself very well to repeated measures of behavior over time.

Most teachers have skills in measuring permanent products. The recording and graphing of permanent products may be a precise and observable technique for evaluating the effectiveness of the teaching-learning process.

Informal evaluation processes, as compared to standardized tests, are unique in a number of ways: (1) teacher involvement in judgment is the essence of such a procedure; (2) immediate feedback enables the teacher to make any necessary instructional adjustments should results indicate a need for such action; (3) this type of evaluation is timely as it focuses on the behavior as it occurs; (4) such procedures are typically simple and brief; (5) the nature of the scheme itself fits into the newer concept of prescriptive teaching whereby more effort is made to take the style of the learner into consideration as well as his present level of functioning; (6) it provides a longitudinal record of pupil progress in a broader range of situations than the more formal assessment techniques do; and (7) such a scheme provides a check on the use of the application of skills and information, learned within the specific confines of a given lesson, to other situations (Brown, 1972, p. 389).

Measurement requires extra time and effort, but this cost is quite minimal in comparison to the benefits that are derived by children as a function of applying evaluation data to classroom instructional procedures (Hall, 1972, p. 403). A very systematic and comprehensive record of pupil progress is accomplished by employing any or several of these techniques, i. e. criterion-referenced measures, observational recording, and measurement of permanent products.

Generally, this discussion of evaluation procedures has been concerned with measuring academic responses. Yet, schools should also be concerned with, and held accountable for, social and emotional growth. Behavior analysis is concerned with behavior whether it is academic, social, or emotional. In fact, the first behavioral analysis in education concentrated upon deviant social behaviors such as excessive crying (Williams, 1959), abnormal gross motor play behaviors (Patterson and Brodsky, 1966), and simple isolate behaviors (Bijou, 1966; Allen, Hart, Buell, Harris and Wolf, 1964). These early emphases led some observers of the field to assume that the products of behavior analysis were only appropriate for young children or persons with grossly deviant behavior. However, there are clear trends to work with normal and older students and emphases on productive behaviors such as appropriate social interactions (Clark, 1972) and creative (Goetz and Baer, 1971; Goetz and Salmonson, 1972) and intellectual responses (Knight, Hasazi and McNeil, 1971; Hamblin, Hathaway and Wodarski, 1971). Certainly the field appears to be perfectly open and the research methods and strategies sufficiently flexible to accommodate work with any behavior which is important in education. The only requirement is that behavior be defined specifically enough to allow for reliable measurement (Ramp and Hopkins, 1972, p.xi).

It is suggested that teachers initiate their evaluation procedures with an academic area because their previous experiences may make it the most comfortable starting point. Most teachers are acquainted with giving and scoring tests, evaluating special reports and projects, and preparing report cards for parents. If classroom behavior is so disruptive that it is useless to attempt to teach academics until some behavior control has been established, the teacher should initiate evaluation procedures in the social skills area. The basic assumptions for suggesting that teachers focus on academic evaluation are: (1) the teacher has sufficient control of student social behavior and (2) the major classroom concerns are instructional.

Two basic procedures have been developed to aid the classroom teacher in scientifically evaluating the effects of his or her classroom procedures. These designs are called "reversal" and "multiple baseline" scientific verification procedures (Baer, Wolf and Risley, 1968). These designs are in contrast to the experimental-group/control-group design which is inappropriate for studying individual behaviors because it involves superimposing a set of population characteristics on an individual child. The reversal and multiple baseline designs provide useful and sound ways to demonstrate causal relationships, and can be applied to single subjects or to single groups (Hall, 1971, p. 22). Hall (1972, p. 403) writes that emphasis in research is shifting from investigation of general questions to investigation of specific procedures and their effects on classroom groups and individual children. The new emphasis centers around observing and recording behavior, which were previously discussed.

The reversal design employs five steps: (1) $Baseline_1$--Scientifically define and record the operant (existing) level of the behavior prior to the institution of teaching conditions; (2) $Condition_1$--Begin teaching procedures while continuing to record the strength of the behavior; (3) $Baseline_2$--Withdraw teaching procedures (return to $Baseline_1$ conditions) and continue to measure the behavior; (4) $Condition_2$--Reinstate the teaching procedures (return to $Condition_1$); (5) Post Check--Periodically measure the behavior beyond the formal end of instruction to see if behavior levels are being maintained.

Two steps are employed in the multiple baseline design: (1) Baseline--Concurrently record several behaviors of one student or group, record the same behavior of several students or groups, or record the same behavior of an individual or group under various stimulus conditions; (2) Conditions--(a) begin teaching conditions with one behavior until a change in that behavior is observed, (b) begin the same procedures with the second behavior, (c) successively begin teaching procedures with the third behavior, etc. (Hall, 1971, pp. 23-24). Both the reversal and multiple baseline designs allow teachers to make probability statements that their instructional procedures are responsible for changes in student behavior.

The self-accountability of teachers is both desirable and possible. New techniques in individualizing instruction, in observing and recording behavior, and in research designs make it possible for each teacher to systematically instruct, and scientifically determine which programs and procedures help children and which are ineffective. Self-accountability, then, is the teacher's critical analysis of the effectiveness of the learning process.

The Effects of Knowledge of Performance

Recording and graphing behavior is basic to self-accountability. By graphing data, the teacher creates a visual representation of behavior which is easily interpreted by students and parents. Not only is this an effective evaluation technique, but knowledge of performance has some interesting effects on the subjects involved.

The phrase "knowledge of performance" or "knowledge of results" at first glance seems self-explanatory and specific. However, it refers to a great variety of environmental changes or stimulus presentations ranging from indicating to the learner that he is correct to providing elaborate, informative, corrective materials when he has made a mistake. "Nonverbal" feedback consequences are as varied. A student might discover how well he has assembled a piece of apparatus by plugging it in and trying it out; he might test a formula he has invented by going into the laboratory and blending the ingredients; he might observe a computer simulation of his patient's vital signs before, during, and after the treatment he has proposed; he might watch a model of the bridge he has designed displayed on a cathode ray tube and undergoing stress from traffic and winds. In fact, any consequence discriminable to the learner and regularly related to his previous performance can be designated as knowledge of results (Geis and Chapman, 1973). This 'feedback' is a method of controlling the output of a system by reinserting into it the results of past performance (Weiner, 1950).

It is sometimes said that being correct, i.e., "knowing you are right," is reinforcing in itself and need not be linked to other reinforcers. This claim for an autonomous status of knowledge of results is particularly common among some educators. They are concerned that the user of other reinforcers, either instead of or in addition to knowledge of results, not only weakens the effect of "the joy of learning for its own sake," but also smacks of immorality and bribery. The purpose of this study is not to discern what particular type of reinforcer knowledge of results may be, but to disclose evidence that knowledge of results may or may not function as a reinforcer.

Most studies are not directly aimed at investigating whether or not knowledge of results is reinforcing. The question usually being attacked is a broader one: Does feedback in some way affect performance during and after that performance (Geis and Chapman, 1973)?

Effects of Feedback on Performance

There is evidence that under some circumstances feedback affects performance.

Hundal (1969) described a study designed to assess the motivational effects of knowledge of performance in a repetitive industrial task. Eighteen male workers were divided randomly into three groups. Group A received no information about their output, Group B were given a rough estimate of their output, and Group C were given accurate information about their output and could check it further by referring to a figure that was placed before them. Hundal found evidence of significant increased output with increase in degree of knowledge of performance.

In an investigation of the relationship between test anxiety and feedback in programmed instruction, Campeau (1968) found that feedback was a significant variable in the performance of fifth-grade school girls on a programmed instruction lesson on earth-sun relationships. Post-instructional test scores were higher for those high-anxiety females who had feedback during learning. Low-anxiety female students who had no feedback had higher post-test scores than high-anxiety females who had no feedback. Male students showed no similar regularity.

In two studies by Anderson, Kulhavy and Andre (1970), 356 subjects in two experiments completed a programmed lesson on the diagnosis of myocardial infarction. The experiments were conducted using a computer-based instructional system which insured that, unlike many previous studies, the subject responded before he received knowledge of the correct response. The subject who received knowledge of correct response after they responded learned significantly more than subjects who received no knowledge of correct response or who could peek at the correct response before they responded.

Using materials that might be considered "programmed," Wittrock and Twelker (1964) found an interesting relationship between knowledge of results and rules. While rules alone proved most effective in teaching subjects to decode ciphered sentences, knowledge of results was especially useful when rules were not supplied. It did not add to teaching effectiveness when supplied in conjunction with rules, supporting the authors' contention that knowledge of results enhances learning retention and transfer when information contained is not greatly redundant.

Madsen, Becker and Thomas (1968) varied systematically the behavior of two elementary school teachers. The study was designed to determine the effects on classroom behavior of rules, ignoring inappropriate behaviors, and showing approval for appropriate behaviors. Following baseline recordings, the rules and the feedback conditions were introduced one at a time. The main conclusions were that: (a) rules alone exerted little effect on classroom behavior, (b) ignoring inappropriate behavior and showing approval for appropriate behavior (in combination) were very effective in achieving better classroom behavior, and (c) showing approval for appropriate behaviors is probably the key to effective classroom management.

Cossairt, Hall, and Hopkins (1973) studied the effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. Feedback increased teacher praise for teacher B and feedback plus social praise significantly increased teacher praise for both teachers A and B. Teacher C received a package condition, that is, experimenter instruction, feedback, and social praise. This teacher's praise also increased significantly. This study featured an examination of the complete chain of behaviors from experimenter through teacher through student. Student attending behavior increased as teacher praise increased.

James (1972) investigated the question: Does providing direct formalized feedback to participants accelerate their learning, thereby producing greater changes in attitudes and behavior? She concluded that the use of a direct formalized feedback technique in group training produced behavioral changes which can be observed by other group participants.

Leiterberg, Agras, Thompson and Wright (1968) used feedback in two behavior modification single-case research studies. A claustrophobic patient and a knife-phobic received graduated practice in facing their phobic stimuli; length of time the claustrophobic patient stayed in a small dark room per trial and length of time the knife-phobic patient kept knife exposed per trial were measured. In both experiments, when feedback of these time scores was withdrawn, ongoing progress was retarded. Reinstatement of feedback led to renewed improvement. In Experiment 2, adding and removing contingent verbal praise against a constant background of precise feedback did not significantly alter rate of progress.

Under certain conditions, then, various presentations of knowledge of results have been shown to: (1) have motivational effects, (2) to be a significant variable in increasing post-test performance when subjects were anxious, (3) to increase achievement in some kinds of computer-based instruction, (4) to be useful when rules are not supplied and sometimes to be more effective than rules, (5) to increase appropriate behaviors in both students and teachers, (6) to produce positive behavior change in group therapy participants, and (7) to decrease phobic behaviors.

No Feedback Effects on Performance

There are also a number of studies relative to the effects of knowledge of results which suggest that feedback concerning performance has no effect on the subjects involved.

Ripple (1963) compared teaching material in a variety of forms including a standard programmed text with and without "reinforcing feedback." The author concluded (from a comparison of criterion test scores of the feedback and no-feedback groups) that there was no differential learning or retention.

An oft-cited study by Feldhusen and Birt (1962) used a short program presented to college students. Nine experimental groups of thirty subjects each received instruction with nine variations of programmed materials to test the hypothesis that there would be no differences in learning among the groups. The topic of the programmed materials was concerned with teaching machines and programmed learning. The authors concluded that the nonconfirmation groups (those who did not receive feedback) did not significantly differ from the confirmation group.

Moore and Smith (1964) also reported no differences on post-test scores between knowledge of results and nonknowledge of results groups of college students who used a version of the Holand-Skinner psychology program (The Analysis of Behavior, New York, McGraw-Hill, 1961). The experimenters tried a variety of feedback conditions (knowledge of results alone, knowledge of results plus pennies, knowledge of results plus light), none of which seemed to affect learning significantly. However, errors within the program were fewer for the knowledge of results group.

While findings which indicate no effect of feedback are disconcerting, even more damaging would be evidence that feedback actually hindered performance. In a study by Swets, Millman, Fletcher and

Green (1962) subjects were taught to identify multidimensional, non-verbal sounds utilizing a computer-based teaching system. Detrimental effects of extensive feedback were demonstrated when the performances of feedback and no-feedback groups were compared.

The results in the literature are conflicting and puzzling. A number of design and interpretation problems may lie behind some of the ambiguity. Some of the variables to be considered are: kinds of feedback, schedules of feedback, method of giving feedback, kind of task, and organismic variables such as age, intelligence, sex, and motivation. These design and interpretation problems will be discussed in detail later in this chapter.

Schedules of Reinforcement

A considerable body of literature has concerned itself with schedules of reinforcement (Ferster and Skinner, 1957) i.e., the effects of arrangements in the provision of feedback. Schedule of reinforcement refers to the determined contingencies of applying reinforcement. Reinforcement can be continuous--every desired response is reinforced, or intermittent--only certain responses are reinforced. The two basic categories of intermittent reinforcement schedules are ratio schedules (contingent on number of responses) and interval schedules (contingent upon passage of time between responses). The number of responses or the required passage of time can be fixed or varied. Human behavior has been shown to be differentially influenced by the various schedules; ratio schedules produce high rates of responding, interval schedules produce low rates of responding, fixed schedules produce pauses in responding, and variable schedules typically produce steady responding (Hall, 1971, pp. 9-13). Reinforcement schedules, then, should be and have been carefully considered in feedback studies.

Two studies (Johnston, Maerteus and Schooley, 1969; Hillman, 1969) have dealt specifically with effects of various knowledge of results conditions on arithmetic performance of elementary school children. In both studies, one group of children received knowledge of results after completing each problem while the other group received knowledge of results after completion of the entire assignment or on the following day. Those receiving per-problem knowledge of results performed significantly higher than the other groups.

Lublin (1965) studied a large group of college students in a programmed psychology course. Students were on schedules of reinforcement including no confirmation, 100 percent confirmation, fixed ratio 50 percent confirmation, and variable ratio 50 percent confirmation. The no-confirmation group scored higher on the criterion test than did those students under the continuous confirmation treatment, and she suggests that "omission of the answers may have caused the subjects to look for confirmation of their responses in succeeding frames." Presumably, this is tied somehow to better attending and, consequently, improved learning. Conversely, the continuous confirmation group may not have engaged in these beneficial searching behaviors and may have learned little from the frames. In addition, Lublin suggests that the post-test more closely resembles the no-answer program, so that the students whose program responses were continuously confirmed may have been handicapped when faced with the post-test.

In a study by Moss and Neidt (1969) the problem was looked at in the context of information theory; both knowledge of results and amount of certainty were varied. University and high school students served as subjects and a short (42-frame) program in insecticides was used. Decrements in learning were found both when items of information were omitted (i.e., lower percentage of knowledge of results) and when uncertainty was reduced. The authors conclude that the effectiveness of knowledge of results is intimately related to the degree of uncertainty; knowledge of results is useful and important when uncertainty is high.

One can also find studies showing no schedule effect. Glaser and Taber (1961) investigated the effects of partial "reinforcement" using a symbolic logic program for high school students. None of the four experimental groups (100 percent confirmation, 50 percent fixed ratio, 50 percent variable ratio and 25 percent variable ratio) differed significantly from each other on the criterion test. The authors suggest that the reinforcing effectiveness of feedback may depend upon the age of the student, specific subject matter, IQ, and probability of correct response.

Krumboltz and Weisman (1962) studied the effect of intermittent confirmation on 121 students by omitting various patterns of affirming answers from a programmed textbook on educational measurement. The schedule included four levels of fixed-ratio confirmation and two of variable-ratio confirmation. No difference was found in criterion test scores for variable vs. fixed ratio schedules, nor from 0 percent, 33 percent, 67 percent, and 100 percent schedules. There also was no interaction effect. However, students made fewer frame errors in programs with more confirmation.

Rosenstock, Moore and Smith (1965) used four different schedules of feedback with a programmed mathematics course: 100 percent, 20 percent fixed, 20 percent variable and 0 percent. Subjects were 96 sixth-grade students enrolled in four mathematics classes. A program in set theory was selected for use. Partial knowledge of results did not seem to enhance learning, but again, fewer program errors occurred under conditions of increased knowledge of results.

The above studies indicate that the effects of different schedules of feedback are apparently affected by a number of other factors such as the level of uncertainty expressed by students, the specific subject matter, student attending behavior, and the age and intelligence of the students involved.

Student Use of Feedback

Since student attributes seem to be important in feedback studies, other writers attended to the student's involvement with feedback. One way to discover the reinforcing effect of knowledge of results would be to make feedback available to the learner and record whether or not he profits from its availability.

The effect of self-recording on the classroom behavior of two eighth-grade students was investigated by Broden, Hall and Mitts (1972). In the first experiment, the eighth-grade girl's in-class study behavior increased when self-recording was instituted. Study behavior decreased when her counselor did not provide recording slips, and then increased once self-recording was reinstated. Teacher praise for study was paired (presented in combination) with self-recording, and eventually, the self-recording was discontinued without significant losses in study behavior. In the second experiment, the number of talk-outs emitted by an eighth-grade boy decreased when self-recording was in effect and increased again when self-recording was discontinued. When self-recording was reinstituted in the final phase, there was a slight, though not significant, decrease in talking-out when compared to baseline. The latter subject was not differentially reinforced with praise or attention, and the authors note that doing so might have increased the effectiveness of the procedures used.

Vastbinder (1972) produced a partial replication of the Broden, Hall and Mitts study. At his counselor's suggestion, a seventh-grade boy engaged in self-recording his study behavior. Study behavior decreased when self-recording was introduced and increased to baseline when teacher praise for study behavior was instituted. At this point, the boy was instructed to set a timer for twenty minutes and to study until it rang. This procedure was carried out on two occasions, and these were added to the study as two more treatment conditions. Study behavior increased to 96 percent and 100 percent under the two timer conditions. Perhaps this latter condition was the most successful with this boy because the feedback (i. e. knowledge of amount of time spent studying) was continuous and visual.

Cummings, Schwab and Rosen (1971) investigated the effect of previous performance on a simple addition task which was found to exert a significant positive impact on the goal level set by 80 subjects. When performance effects were accounted for, knowledge of results also influenced goals significantly.

Melching (1966) used a 364-frame linear program on magnetism with 17 enlisted Army personnel. All answers had been deleted from the program, which was then administered individually to each subject. The experimenter, sitting opposite the subject, provided feedback (the printed answer) after each response upon request from the subject. The learner requested feedback on about one-quarter of the frames. Percent requests by subjects ranged from 57 percent to 6 percent. The first conclusion, then, is that subjects tended to request feedback much more often when they were wrong than when they had correctly responded. Low ability students (as defined by scores on a measure of "general intellectual ability") requested feedback about three times as often as high ability students. They also made about three times as many errors in the program.

A series of studies by Geis, Jacobs, Spencer and Neilson (1970) strongly supports Melching's findings. Several different programs were used with college students as subjects. The answer to each frame was either on the back of a card containing the frame or, in some instances, was exposed when the subject removed a piece of masking tape. It was consistently found that: (1) on the average, students checked far less than 100 percent of the answers; (2) students varied among themselves in regard to the percent of answers they checked (data indicate that each subject was consistent in his checking rate over a variety of programs); (3) clear, significant and positive correlations were obtained between erroneous responding and checking. Thus, though the checking rates differed widely from student to student, with only a few exceptions, the probability of checking, regardless of base rate, is higher after a student has emitted an erroneous response than when he has been correct.

Researchers seemingly agree that student use of feedback is influenced by a number of conditions. Some of these are: (1) ability level of students; (2) type of task; (3) fear of knowledge that their response was wrong; and (4) kind of feedback.

Teacher Use of Feedback

Another group of researchers chose to study the effects of teachers' use of feedback. Feedback was often used in an attempt to improve teacher training programs or to improve the classroom teacher's performance.

Birch (1969) considered the effects of having student teachers view video tapes of their own teaching performance. The key element in Birch's data was the suggestion that categorizing and recording the frequency of one's own behaviors may be a powerful procedure leading to changes in the recorded behaviors.

Thomas (1971) used a similar procedure in an attempt to increase teaching behaviors, such as the delivery of tokens and the use of praise. Video tapes were prepared for use in training the teachers to count and graph specific classes of behavior. Each training tape provided a verbal definition of one or more categories of teacher behavior, instructions on how to record and graph the behaviors being observed, examples of the behaviors, and a five minute teaching sequence for practice in recording. The available data suggest that it is possible to obtain increases in the frequency of token deliveries and behavior-specific praise statements by having the teacher analyze video tapes of her own teaching.

Rule (1972) modified three teaching behaviors--praise, on-task contacts, and off-task contacts--in nine teachers. A multiple baseline design was employed to test the effects of three procedures: instructions and experimenter feedback, video tape scoring of one's own behavior, and a direct intervention procedure in which the experimenter temporarily replaced the teacher whose teaching behavior fell below criterion. Direct intervention was most effective in changing teachers' behavior. Smaller changes in the subjects' rates of praise, on-task contacts, and off-task contacts occurred during the video-scoring procedure, and no predictable change occurred on the instruction plus feedback condition.

Bapst (1972) studied the effect of systematic student response upon teaching behavior. The instructors asked students to continually indicate how well they felt they were understanding the material being presented. The hypothesis that teachers would teach better if they received feedback through an in-class student response system was supported.

Novak and Moser (1972) studied the effect of timed pupil feedback on the teaching behaviors of biological science teachers. It was found that students were able to give effective feedback to teachers and that this feedback caused changes in teacher behavior. Pupil's responses indicated a positive correlation between the amount of time spent lecturing and the percentage of students responding to "understand" on rating sheets; a negative relationship existed between understanding and discussion. Teachers receiving this feedback either maintained or increased amount of lecture during and after the treatment month.

Implications

There are a variety of variables to be considered which confound experiments in which knowledge of performance is being manipulated. They will be discussed in the following section.

Permanent or momentary organism-centered variables (e. g., anxiety, IQ, sex, age achievement motive) might confound experiments. Extensive research on one such variable was conducted by Campeau (1968) whose work was reviewed earlier. A major variable in her research has been anxiety, specifically test anxiety as measured by the Test Anxiety Scale for Children. Accentuating the test-like features of the situation by omitting answers in the program should, the author contends, adversely affect high-anxiety subjects. Furthermore, a comparison of high- and low-anxiety subjects under feedback conditions should reveal higher achievement scores for the high-anxiety group (implying, in the context of this paper, that feedback may be reinforcing, or at least more reinforcing, to anxious students). The differences Campeau found in her studies were not significant for the male subjects. High-anxiety girls did somewhat better, but not significantly better on immediate post-tests than did low-anxiety girls when both groups had received feedback in the program. High-anxiety girls in the feedback group showed quite dramatically better gain scores than those in the no-feedback group.

A number of writers (e.g., see Taber, et al., 1965) suggest that factors such as age, motivation, and IQ may well interact with feedback. However, little research has been directly aimed at investigating such relationships.

A second set of variables which might affect the status of feedback involves the task itself: the kind of task (e.g., learning rote material in which the components have little interaction with each other vs. learning concepts which are related) or task complexity. Feedback might be more reinforcing when one is executing a complex motor coordination task than when he is merely recognizing a correct item in a choice situation. This may be related to the idea that the probability of error interacts with the reinforcing effect of feedback since the chance of making some error is usually increased when a chain of responses is called for, as in the case of production.

A third area might be called "kinds of feedback." Certain types of feedback may, in an absolute sense, be more reinforcing than others or may be more reinforcing under certain conditions. For example, given a two-choice discrimination task in a program frame, the student might be reinforced by an indication of "correct response" in the form of a light going on after he emits the right answer. Failure to produce the "correct" signal is logically equivalent to producing an "incorrect" signal, in the two-choice situation.

At the other extreme might be a complex motor task. When learning to pronounce a French word or to write a Thai symbol, binary feedback limited to "correct"--"not correct" may prove unreinforcing (Geis and Chapman, 1972).

"Kind of feedback" might also refer to varying conditions for the same feedback, i.e., the same information. For example, in a study by Anderson, Kulhavy and Andre (1970), one group received answers only after they had made an incorrect response while other students received feedback only after they had made correct responses.

Lumsdaine (1961) suggests that if knowledge of results functions as a reinforcer, it is likely to do so under special conditions. One of these involves low probability of correct response (or, conversely, high probability of error).

Melaragno (1960) investigated negative feedback. He concluded that some spaced negative reinforcement does not impair learning, but that massed negative reinforcement seems to do so.

MacPherson, Dees and Grindley (1948-1949) proposed that the important function of knowledge of results varies with the stage of learning. In the initial stages of learning, they contend, the directive or informational effect of knowledge of results is important; after performance has stabilized, the "incentive" function of knowledge of results assumes greater importance. Finally, when proficiency has reached a high level, overt formal knowledge of results seems to be of little value.

Another suggestion relates item difficulty and the reinforcing effect of confirmation. Holland (1965) after a review of the literature, points out that no difference between confirmation and nonconfirmation is found when programs with low error rates are used. On the other hand, studies which utilized high error rate materials tended to show an advantage in the confirmation group. In such programs, students' confidence in their own answers would not be great. Analysis of student behavior and knowledge of results suggest that knowledge of performance may be a reinforcer when uncertainty or probability of emitting an incorrect response is high, or where confidence is low (Geis and Chapman, 1972).

CHAPTER III

PROCEDURE AND METHODOLOGY

The purpose of this study was to investigate the effects of knowledge of student performance in arithmetic on both the teachers' and students' behavior.

Subjects and Setting

The study was conducted in one elementary and two junior high schools in a community in Southwest Ohio; population--37,500; median income--$14,253 (Economic Profile: Fairborn Chamber of Commerce, 1973-74). The Superintendent of schools in this community gave permission to the investigator to seek volunteers for teacher-subjects in this study. All elementary and junior high school teachers of educable mentally retarded students in the school system were given the opportunity to volunteer. Three teachers of students classified as educable mentally retarded (IQ 50-80 as measured by the Revised Edition of the Stanford-Binet Intelligence Test; range required by the Ohio Department of Education) were selected by using simple random sampling. Names of volunteers were coded and a table of random numbers was used to select the participants. The population was limited to educable mentally retarded students and their teachers. This was done because results of studies indicate that knowledge of results may be a reinforcer when uncertainty is high, or where confidence is low (Geis and Chapman, 1972). The behaviors of the teachers were also studied because results of research indicate that knowledge of student performance changes teacher behavior (Bapst, 1972; Novak and Moser, 1972). A delimitation of the study was that educable mentally retarded students and their teachers were selected from one particular community due to constraints of time, money, and availability.

Two junior high school teachers and one elementary school teacher were randomly selected to participate in the investigation. Each teacher selected three students who had low achievement in arithmetic (see Operational Definition, p. 10). One junior high group was designated the experimental group; the elementary group became the replication group, and the second junior high group served as the control group. Table 1 describes the student-subjects.

Teachers were told that they were volunteering to participate in a study of the types of arithmetic errors that educable mentally retarded children make. The teachers were involved in supplying each student with feedback (knowledge of arithmetic performance in daily quizzes as recorded on a graph). However, they had no knowledge of the hypotheses either before or after the experimental conditions were

TABLE 1

EMR STUDENTS SELECTED AS LOW ACHIEVERS IN ARITHMETIC

N = 3 Group	N = 9 Subjects	Chronological Age	Grade Placement Arithmetic (WRAT)	Intelligence Quotient
Experimental	A_1	14-1	3.9	73
Experimental	A_2	13-7	3.9	68
Experimental	A_3	14-7	1.9	64
Replication	B_1	7-11	1.4	56
Replication	B_2	9-9	kg. 1	66
Replication	B_3	7-4	kg. 5	67
Control	C_1	13-2	1.5	74
Control	C_2	13-5	3.4	67
Control	C_3	14-7	4.4	69

implemented. Teachers did receive a written report at the completion of the study on the information they thought they were gathering for the investigator; that is, the types of arithmetic errors their students were making.

One criterion the participating teachers had to meet was the ability to maintain classroom control. This criterion was established by the investigator so that disruptive behavior, either by the selected student or his classmates, was not a confounding variable in the student's low arithmetic achievement. The ability to maintain classroom control was determined by considering the opinions of the building principal and the teacher him- or herself in regard to this matter.

Selection of Dependent Variables

The investigator selected number of correct arithmetic responses by students and the amount of time teachers spent in arithmetic lessons as the dependent variables. Each dependent variable will be discussed in the following section.

Number of Correct Arithmetic Responses

The arithmetic performance of educable mentally retarded students was selected as a dependent variable for several reasons. The first consideration was ease in measuring output. Arithmetic computation results in a permanent product. Such products are tangible and can be measured subsequent to the student's behavior

(Cooper, 1974). They provide concrete information for informal evaluation by teachers. Moreover, most teachers already have skills in measuring permanent products. This procedure did not necessitate additional training in measurement techniques for the teacher-participants.

A second consideration was the social significance of possessing arithmetic skills. Such skills are basic to occupational competence--one of the major goals in educating mentally retarded persons (Kirk, 1962, p. 112). While the same case can be presented for selecting reading achievement, the investigator determined that the reading programs involved a larger number of variables which could affect daily quiz scores.

Amount of Time the Teacher Spent in Arithmetic Lessons

The investigator assumed that spending more time in arithmetic lessons was an indication of increased effort by the teacher to improve student performance. Also, amount of time spent in arithmetic lessons was selected because this was a variable which could be measured mechanically (by tape recording sessions). Tapes and tape recorders were readily available to the teachers. Tape recording sessions provided an accurate, permanent record of teacher behavior.

Another concern was that the presence of an observer might indicate to teachers that teaching behavior was being measured. Teachers were told that the taped arithmetic sessions were analyzed to determine the types of questions the students asked. In addition, the necessary training and honorariums were further constraints in using classroom observers.

Experimental Procedures

Each child selected by the teachers was assessed by the investigator utilizing the Science Research Association Arithmetic Series Diagnostic Tests (Montgomery County Board of Education, 1965) and the Wide Range Achievement Test (WRAT) (Guidance Associates, 1965). An individualized new math program was planned for each child by the investigator in cooperation with the child's teacher. The new math program was based on the Science Research Association Arithmetic Series. In an effort to equalize the degree of difficulty of materials assigned to each child in the study, each daily quiz had ten problems. Five of the problems in each daily arithmetic assignment were problems that each student knew how to compute, as revealed by the initial assessment and the continuing daily assessments. The other five problems required arithmetic skills which the student was taught in that day's instruction. For each child, the arithmetic skills were taught sequentially as presented in the Science Research Association Arithmetic Series.

Teachers A, B and C were instructed to teach arithmetic in their "usual" manner. Individual teachers may have used other reinforcers, either systematically or unsystematically. Examples of the types of reinforcers used might include verbal praise, smiling, candy, "stickers," working problems for students, and a variety of other reinforcers. The differential effects of these other reinforcers were controlled by collecting baseline data on individual teachers.

The teachers were asked to tape-record the daily arithmetic lessons; a lesson being from the time the teacher changed from the previous activity to teaching arithmetic until she changed to the next activity. This taped session sometimes included periods of silence when the students were studying or taking a test. However, the teachers were instructed to continue tape recording until an activity other than arithmetic was introduced. The amount of time spent in daily arithmetic instruction was calculated from the time the teacher began recording until she stopped recording. To determine observers reliability, a second observer listened to the tape and calculated the duration of the experiment.

All daily quizzes were given to the investigator during weekly visits. This procedure not only verified the teacher's records of student performance, but lent supporting evidence to the teacher's assumption that the investigator was interested in studying the types of errors students made.

Research Design

The basic design paradigm for the present investigation was the multiple baseline technique (Baer, Wolf and Risley, 1968; Hall, 1971) utilizing multiple subjects (multiple teachers and multiple students). The logic of the multiple baseline design is to determine operations that relate functionally to the performance of behavior. Control over a behavior is demonstrated if the behavior can be altered at will by altering the experimental variables. The effects of the experimental operations may be immediately observed on response rates (Kazdin, 1973). The investigator recognized several weaknesses of the single organism design regarding sources of invalidity (Campbell and Stanley, 1963, pp. 6-7). To strengthen the design, a replication group and a control group were incorporated into the study. Teacher B and her students B_1, B_2 and B_3 served as the replication group. Teacher C and students C_1, C_2 and C_3 comprised the control group.

Experimental Conditions

Baseline Conditions: Baseline conditions for all three teachers consisted of (1) tape recording their arithmetic lessons and (2) collecting their students' arithmetic quizzes. This was done to determine the operant levels of each of the behaviors to be studied before instituting experimental conditions (Hall, 1971, p. 8). Tape recording lessons and collecting daily arithmetic quizzes continued throughout the study.

Baselines for teachers A, B and C and students A_1 and B_1 were collected until behaviors reached stable rates (see Operational Definitions, p. 10) or until the response rates moved in the direction opposite that expected during the treatment phases. Baselines of students A_2 and B_2 and students A_3 and B_3 continued until behavior reached stable rates and experimental condition one was introduced to them.

As soon as baselines were concluded for all of the teachers and students, the new math program was introduced into the curriculum. Teacher C continued to tape record lessons and collect daily quiz papers. No other conditions were imposed upon Teacher C and her students throughout the remainder of the study. When stable rates were established for students A_1 and B_1, experimental condition one was introduced to them.

Experimental Condition One: Recording, Graphing, and Sharing Student Performance with the Students: Teachers A and B were instructed to prepare an individual graph for each student. They were instructed to label the horizontal axis "sessions" and the vertical axis "number of correct responses." Numbers were affixed appropriately. Teachers A and B, working individually, explained the graphs to students A_1 and B_1. At the end of each class session, until behaviors reached stable rates, each teacher and her/his student recorded the student's correct responses onto the graph paper. At the end of this phase for students A_1 and B_1, teachers A and B explained the graphs to students A_2 and B_2. Until behaviors reached stable rates, these students recorded their correct responses on the graph paper. The same procedure was followed with students A_3 and B_3.

Experimental Condition Two: Recording, Graphing, Sharing, and Public Display of Student Arithmetic Performance. Teachers A and B cleared a section of their classroom bulletin boards to provide space for the public display of students A_1 and B_1's performance graphs, continuing to add data daily. This began after behaviors reached stable rates in the experimental condition one phase and continued throughout the remainder of the study. The same procedure was followed with the performance graphs of students A_2 and B_2 and students A_3 and B_3. This experimental condition was implemented when stable behavior had been reached in the previous phase. The study ended when stable rates had been reached by all subjects in Group A and Group B.

All the students in the three EMR classrooms had individualized math programs before the study was initiated. Other students accepted any special procedures with the student-subjects as part of an individualized program. None of the students-subjects objected to the public display of their graphs.

Measuring Subject Performance

Teacher Performance

The investigator and the second observer independently calculated the duration of the arithmetic lessons. Each observer played the tape sessions and began timing from the time the teacher changed from the previous activity to the arithmetic lesson until she introduced the next activity. Since teachers were instructed to record only this segment of their teaching, ideally the observers were timing the total activity recorded by the teacher. Practice sessions were held in which both the first and second observer listened to, and timed, "dummy" sessions. This was to insure that both observers were using the same frame of reference to determine the beginning and end of the sessions. It was previously decided that when they had 80 percent agreement on the amount of time teachers spent in arithmetic lessons, they would begin independently calculating the duration of the actual arithmetic lessons. Agreement was determined by dividing the record of the observer with the lower figure by the record of the observer with the higher figure. The quotient was multiplied by 100 and the resulting figure was the percentage of agreement between the records (Hall, 1971, p. 18).

$$\frac{\text{record of observer with lower figure}}{\text{record of observer with higher figure}} \times 100 = \%\text{ of agreement}$$

Interobserver reliability never fell below an average of 90 percent agreement. The time each teacher spent in daily arithmetic lessons during baseline and the two conditions were graphed by the investigator and verified by the second observer. Teachers did not have knowledge of this data.

The amount of time the teacher spent in arithmetic lessons during baseline and during experimental conditions one and two are expressed as mean numbers. The total number of minutes spent under each condition was summed and divided by the number of sessions in that phase of the study. Mean scores were determined for each subject in each phase of the study because the Friedman Test requires a single score in each row by column cell. Friedman was the nonparametric technique selected to treat the data.

Student Performance. The arithmetic performance of each student during each phase of the study was calculated as a mean number. The number of correct responses (as recorded on their graphs) was summed and divided by the total number of sessions during that phase.

Performance of the Control Group. The arithmetic performance of each student during baseline and during the remainder of the study were calculated separately. The number of correct responses on daily quizzes were summed and divided by the total number of sessions during that phase.

The amount of time Teacher C spent in arithmetic lessons during baseline and during the remainder of the study were calculated separately. The total number of minutes spent during each phase was summed and divided by the number of sessions in that phase of the study.

Statistical Treatment

Data was treated three ways:

Subjects' Responses Compared to Themselves. The same behavior of six student-subjects was measured across baseline and two treatment conditions. A change in the level of that behavior for each subject was compared with his baseline response level, and/or with the response level of the previous condition. A statement of causality cannot be made, but an indication of successful change in the subject's behavior can be given. A successful change in this study could be the student's maintained increase in number of correct responses on daily arithmetic quizzes.

The behaviors of the three teacher-subjects were similarly measured. Individual responses were compared across baseline and the two experimental conditions. In this design, each subject serves as his/her own control.

Subjects Compared Within Each Multiple Baseline Group. In the multiple baseline design across individuals used in this study, baseline data were gathered for one behavior across several persons. After behavior stabilized across subjects, the experimental condition was invoked for one subject while baseline conditions were continued for the other subjects. As the experimental condition was extended to include separate individuals, the response frequency was expected to change. This was intended to demonstrate that the behavior of each subject did not change until he was included in the experimental condition. Significance in this comparison need not be ascertained by parametric or non-parametric statistics. Significance in this context can refer to a comparison between the accomplished correction and the level necessary for adequate functioning in our society; it does not refer to a level of confidence that a change has occurred relative to control conditions or groups (Risley, 1969). Social considerations are involved in evaluating the magnitude of improvement brought by the behavior modification procedure. The extent to which the socially desired level of behavioral change is approximated determines the "significance" of the behavioral application.

The Friedman Test (see next section) was applied to the data of Groups A and B as a second way of expressing the results of the data. In addition the Wilcoxon Signed Rank Test (Hollander and Wolfe, 1973, p. 27) was applied to the data of the student-subjects in the control group to determine the effects, if any, of the New Math Program alone on student performance. The Wilcoxon Signed Rank Test is appropriate to use when the data consists of two observations on each of n subjects, and the population is distribution-free (Hollander and Wolfe, 1973, p. 27).

Subjects Compared Across the Three Groups

Since the N was small (N = 9, student-subjects; N = 3, teacher-subjects) and the population was not normal, the use of a nonparametric method to treat the data was advised (Kerlinger, 1964, p. 286; McNemar, 1969, p. 430; Ferguson, 1971, p. 139; Hollander and Wolfe, 1973, p. 1). The data consisted of a set of three observations for a sample of nine individuals in the case of the student-subjects, and a set of three observations for a sample of three individuals in the case of the teacher-subjects. The appropriate statistical treatment then (Kerlinger, 1964, pp. 290-297; McNemar, 1969, p. 434; Ferguson, 1971, pp. 333-335; Hollander and Wolfe, 1973, pp. 139-146) was the Friedman two-way analysis of variance by ranks (1937). The Friedman Test is a nonparametric method for the mixed model situation where the columns stand for C experimental conditions and the rows stand for R individuals (McNemar, 1969, p. 434). Table 2 is a model of the Friedman Test using data from Group A. This statistical procedure was applied to the data of Teachers A and B on their students (Groups A and B) separately. When a null hypothesis was not found tenable, a multiple comparison procedure was applied to determine which treatments differed significantly (Hollander and Wolfe, 1973, pp. 151-154).

TABLE 2

A MODEL OF THE FRIEDMAN TEST

Subjects	Experimental Conditions		
	Baseline	Record, Graph Share Data	Public Display of Data
1	6.00 (1)	7.40 (2)	8.26 (3)
2	5.25 (1)	7.00 (2)	8.33 (3)
3	4.27 (1)	7.00 (2)	7.25 (3)
	$R_1 = 3$	$R_2 = 6$	$R_3 = 9$

CHAPTER IV

RESULTS OF THE STUDY

In this study an attempt was made to determine the effects of knowledge of student arithmetic performance on the behaviors of six educable mentally retarded students and their three teachers. Number of correct responses on the students' daily arithmetic quizzes was measured. The amount of time the teachers spent in arithmetic lessons was measured concurrently with student behavior. The behaviors of a control group were also measured. This latter group consisted of three educable mentally retarded students and their teacher. Data collected in the study are presented in this chapter.

Chapter IV is divided into two main parts; the first part consists of results and the second part is a restatement of the hypotheses. The first part is subdivided into sections which are based on: comparison of each individual's behavior across treatment phases; comparison of subjects within groups; comparison of subjects across groups.

Two baselines are recorded in the subjects' graphs. During $Baseline_1$, teachers were requested to teach arithmetic in their "usual manner" and to give students' worksheets or quizzes to the investigator. Two facts can be ascertained from the $Baseline_1$ data: (1) students were not working math problems every day; and (2) baseline responses fluctuated so greatly that, in a two weeks period, some students had not established stable rates (see Operational Definitions, p. 10). At the end of two weeks of data collecting, all of the students and teachers had Easter vacation. Because some response rates were unstable and no data were collected on any of the subjects for a week, the decision was made that the New Math Program should be introduced to each subject and to record a second baseline until the responses of at least one subject in each of the three groups became stabilized.

Raw data of student-subjects is summarized in Table 3. Raw data of teacher-subjects is summarized in Table 4. The following presentation of results begins with the comparison of the subjects to themselves.

Subjects' Responses Compared to Themselves

Results

In the first section, the behavior of nine students and three teachers is presented. In the multiple baseline research design used in this study, each subject serves as his own control. Behaviors exhibited during the experimental conditions are compared to each other and to the individual baseline data.

TABLE 3

MEAN NUMBER OF CORRECT ARITHMETIC RESPONSES OF EMR STUDENTS

Subject	$Baseline_1$	New Math Program $Baseline_2$	Record, Graph Share Data	Public Display of Data
A_1	7.00	6.00	7.40	8.26
A_2	2.75	5.25	7.00	8.33
A_3	3.00	4.27	7.00	7.25
B_1	9.66	7.83	10.00	9.81
B_2	7.57	7.00	8.50	8.77
B_3	3.71	5.30	5.40	7.00
C_1	5.00	6.04	6.04	6.04
C_2	9.33	6.65	6.65	6.65
C_3	7.20	7.68	7.68	7.68

TABLE 4

MEAN NUMBER OF MINUTES SPENT IN ARITHMETIC LESSONS
BY EMR TEACHERS

Subject	$Baseline_1$	New Math Program $Baseline_2$	Record, Graph Share Data	Public Display of Data
A	39.75	51.00	51.50	51.00
B	34.00	28.33	32.33	30.38
C	26.00	41.30	41.30	41.30

Student-Subject A_1. Subject A_1 is a junior high school female, age fourteen years, one month, and a member of the experimental group. A_1's number of correct responses decreased when the New Math Program was introduced; returned to slightly above $Baseline_1$ when knowledge of performance was recorded, graphed and shared with the student; and increased further when the graphed data was pubicly displayed on the classroom bulletin board. Figure 1 graphically displays Subjects A_1's behavior.

Student-Subject A_2. Subject A_2 is a junior high school boy, age thirteen years, seven months, and a member of the experimental group. Number of correct responses in arithmetic continually increased from $Baseline_1$, through the New Math Program, the Recording, Graphing and Sharing of Data phase, and the Public Display of Data on the Classroom Bulletin Board. A_2's behavior is graphed in Figure 2.

Student-Subject A_3. A_3 is a junior high school girl, age fourteen years, seven months, and the third member of the experimental group. A_3's number of correct responses in arithmetic increased over baseline when the New Math Program was introduced; increased again during the phase of Recording, Graphing and Sharing Data with the student and increased very slightly during the final phase--the Public Display of Data on the Classroom Bulletin Board. Figure 3 represents A_3's responses.

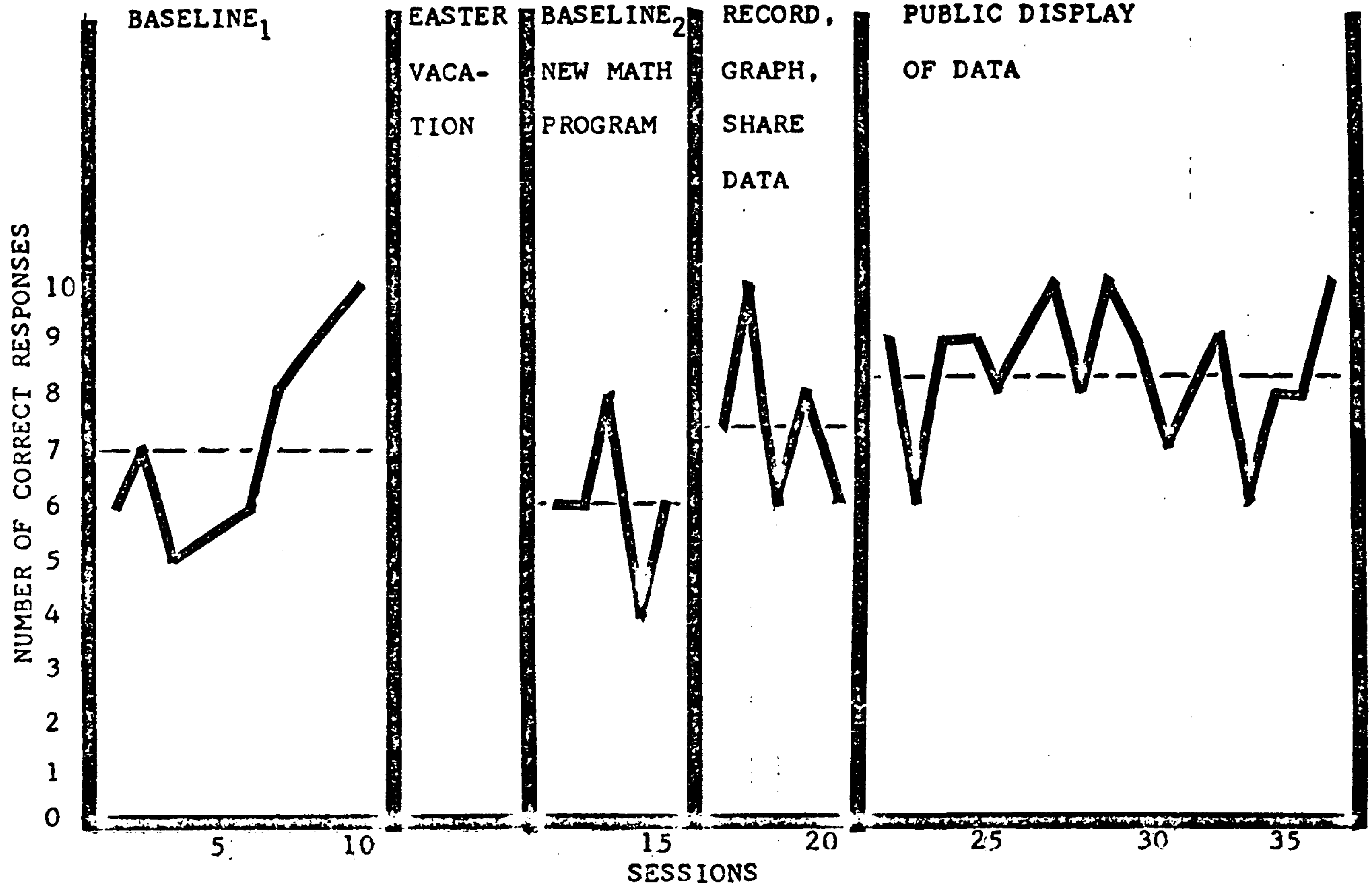

Fig. 1 Student-Subject A_1: Number of Correct Arithmetic Responses

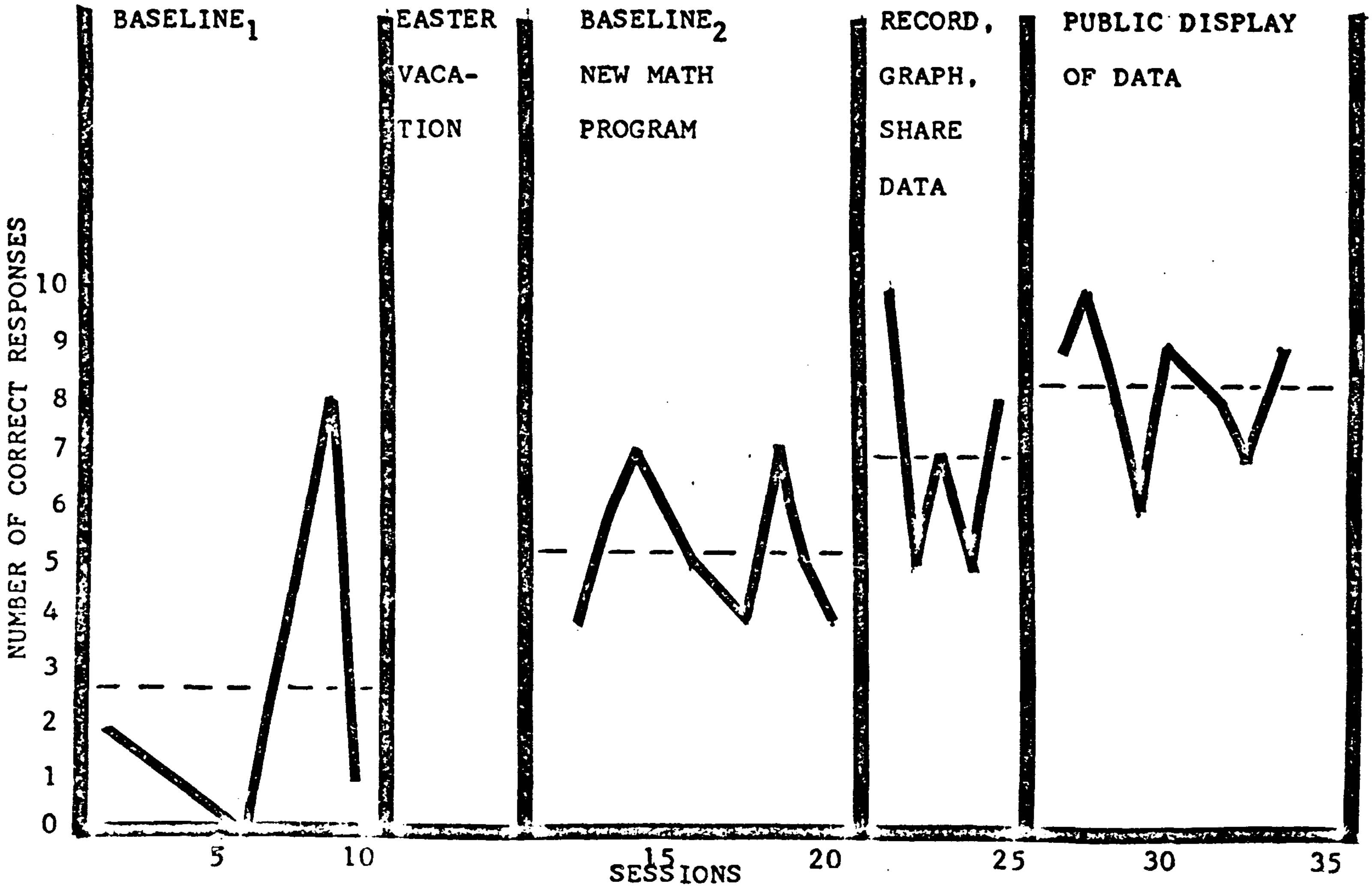

Fig. 2 Student-Subject A_2: Number of Correct Arithmetic Responses

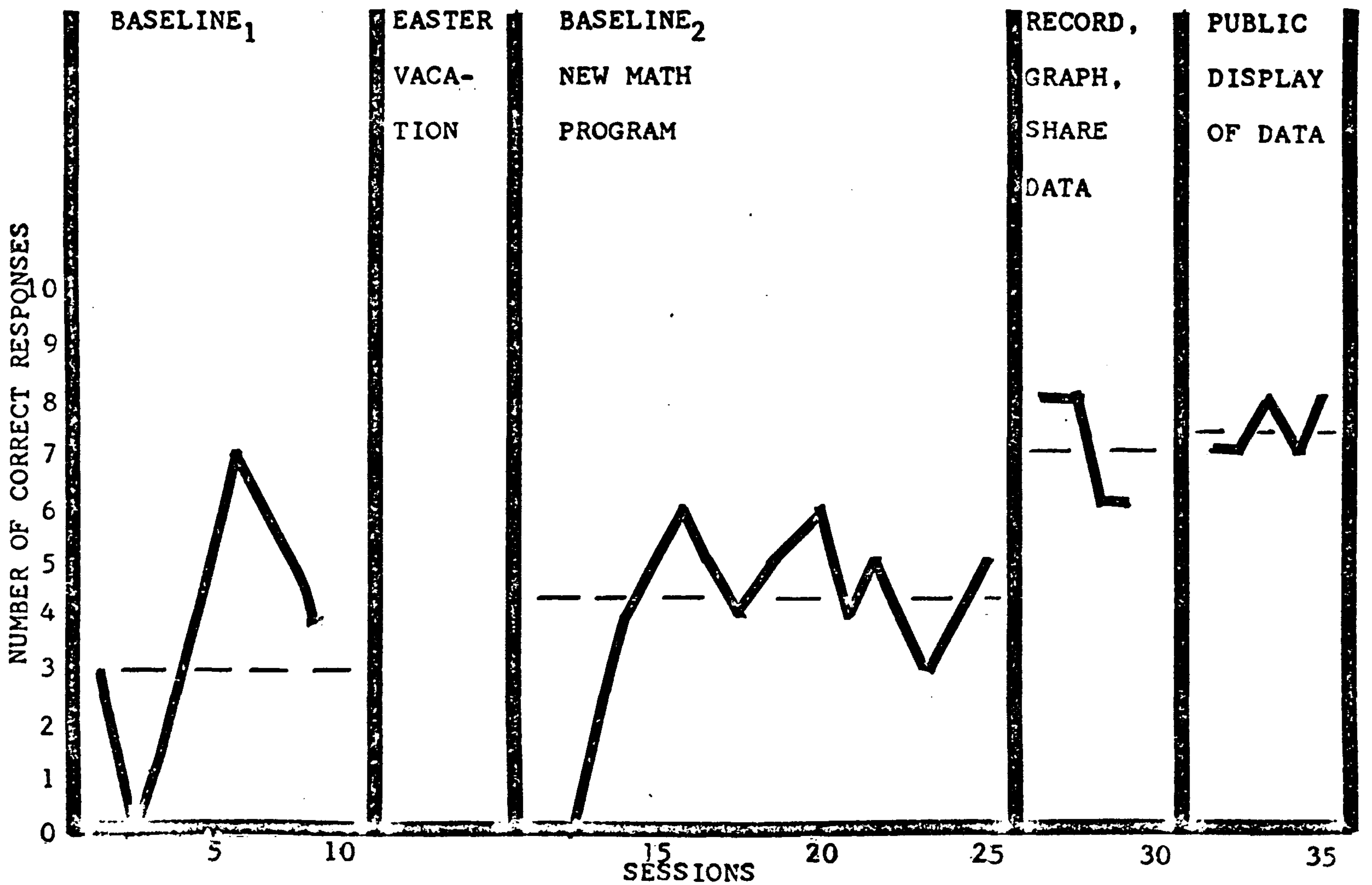

Fig. 3 Student-Subject A_3: Number of Correct Arithmetic Responses

Student-Subject B_1. Subject B_1 is a student in a primary class for educable mentally retarded children. He is seven years, eleven months old and is a member of the replication group. B_1 was the first student in his class to receive the experimental phases. Mean number of correct arithmetic responses during baseline was high for B_1. Quiz papers during this phase reflect B_1's knowledge of number facts up to ten. Subtraction was introduced in B_1's New Math Program. Mean number of correct responses is lower during the New Math Program phase of the study and increases over both Baseline_1 and the New Math Program phase when the experimental conditions are introduced. Figure 4 presents B_1's behavior during the study.

Student-Subject B_2. B_2 is a female student, nine years, nine months old, and a member of the replication group. Mean number of correct responses decreased from baseline rate during the New Math Program phase. However, mean scores increased during the first experimental condition and increased further during the second experimental condition. Figure 5 graphically presents B_2's behavior.

Student Subject B_3. B_3 is a female, seven years, two months old, and the third member of the replication group. This subject's Baseline_1 behavior shows a steady decline in number of correct responses in arithmetic. Mean scores during the New Math Program phase and the Recording, Graphing and Sharing Data phase increased somewhat from Baseline_1, with a more noticeable increase during the

Public Display of Data on the Classroom Bulletin Board phase. B_3's behavior is graphed in Figure 6.

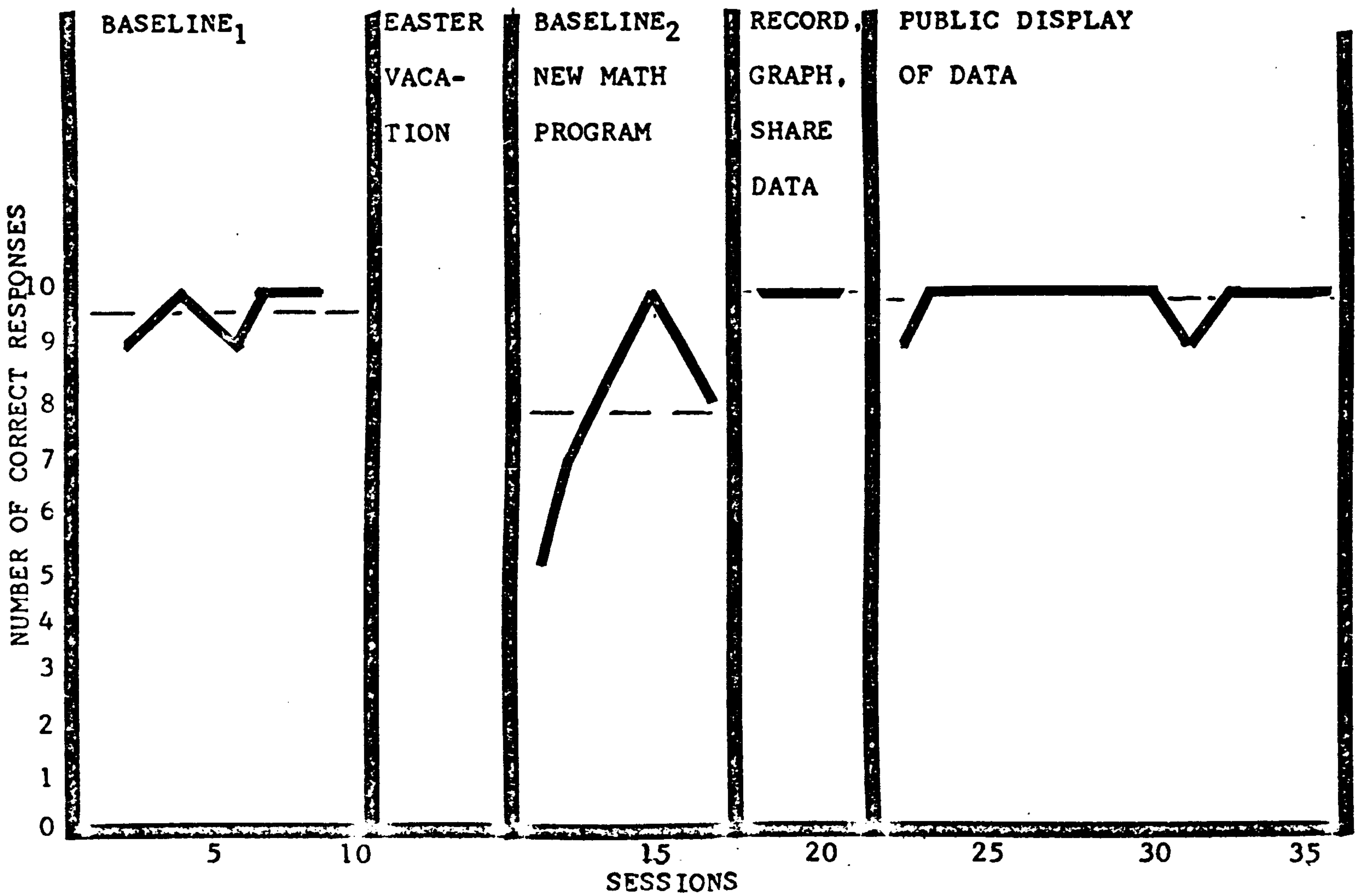

Fig. 4 Student-Subject B_1 : Number of Correct Arithmetic Responses

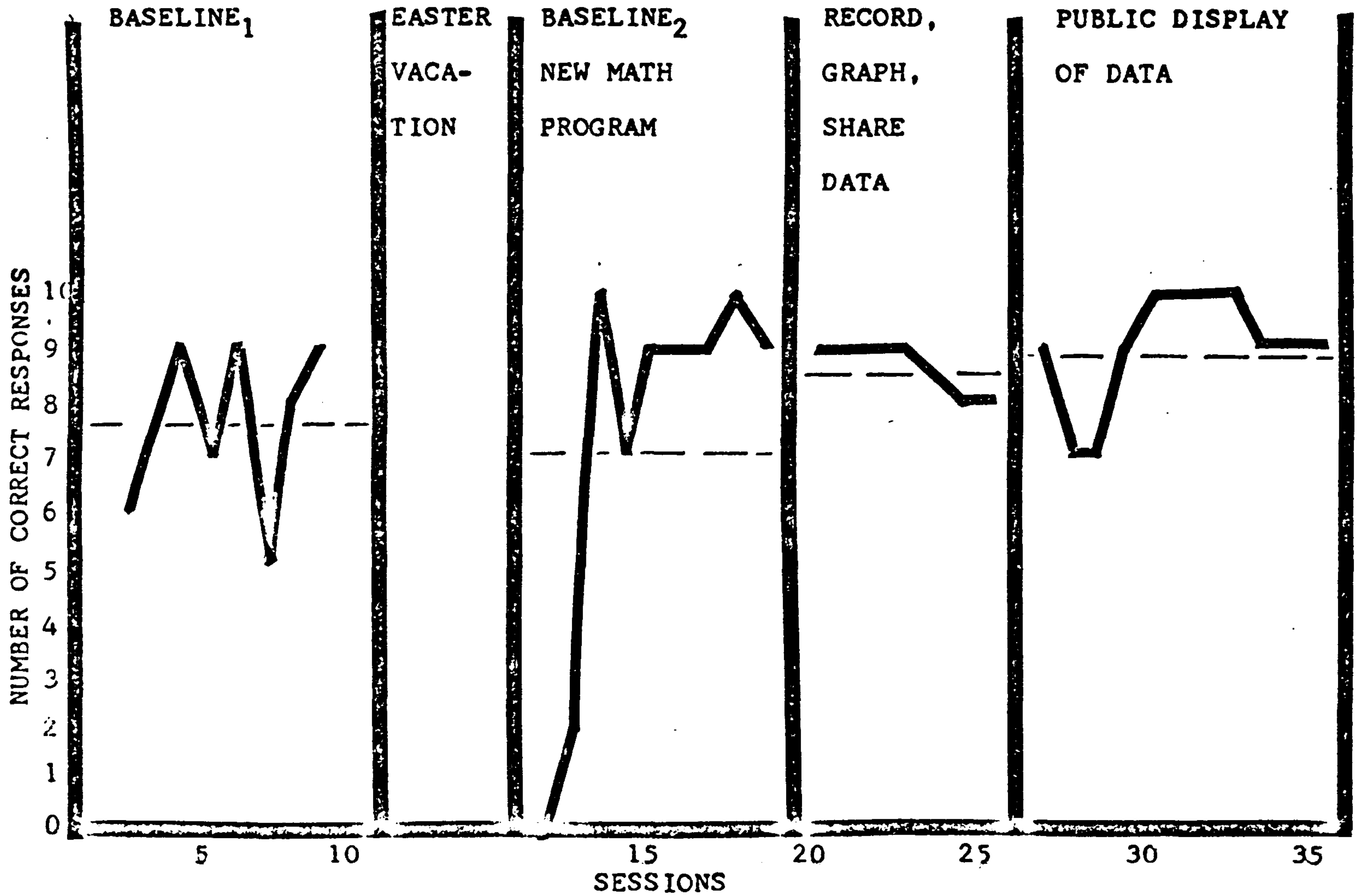

Fig. 5 Student-Subject B_2 : Number of Correct Arithmetic Responses

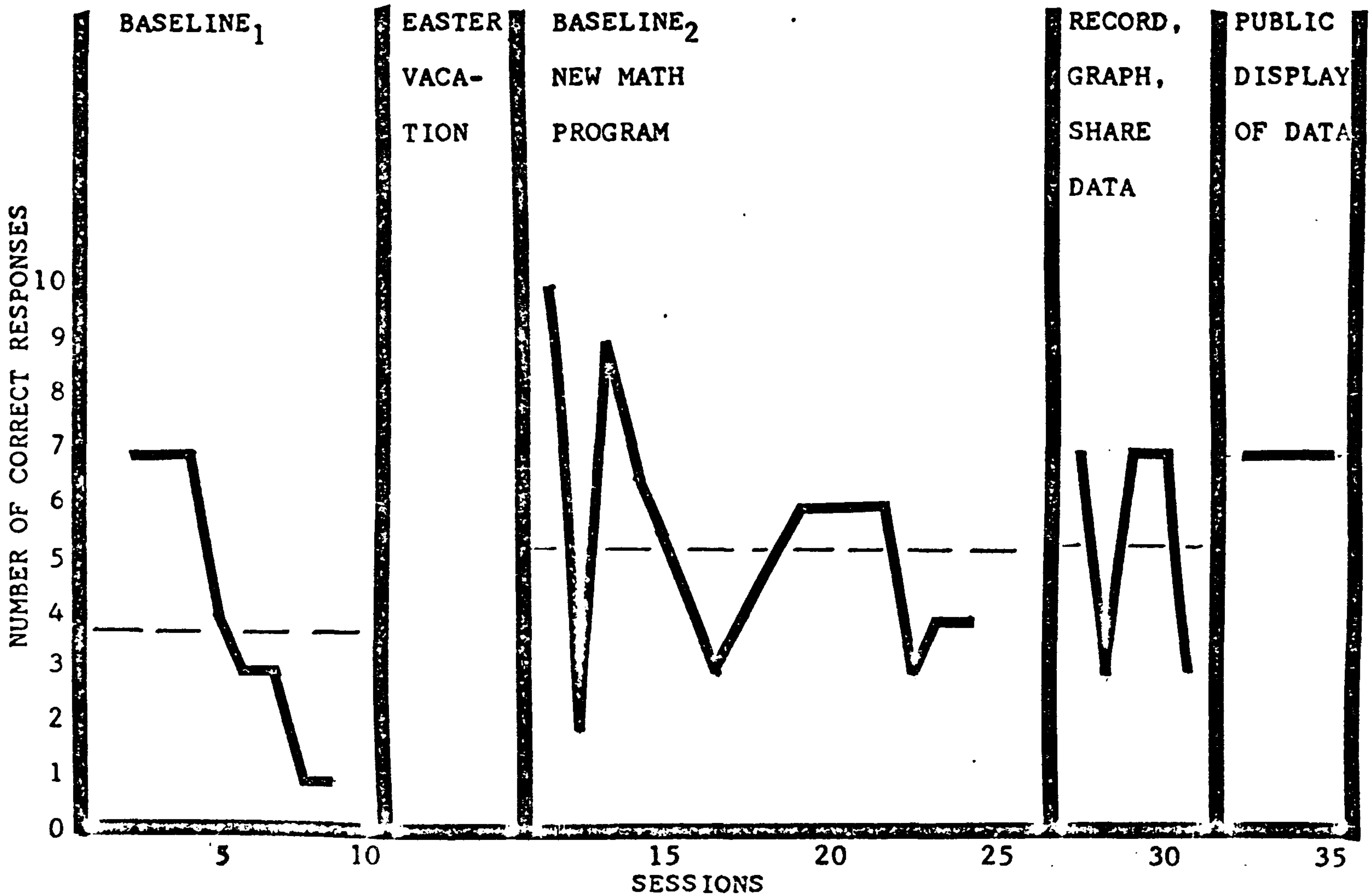

Fig. 6 Student-Subject B_3 : Number of Correct Arithmetic Responses

Student-Subject C_1. This subject is a junior high school female, age thirteen years, two months, and a member of the control group. C_1's number of correct arithmetic responses increased from $Baseline_1$ during the New Math Program phase. C_1's responses are expressed in Figure 7.

Student-Subject C_2. C_2 is a female, thirteen years, five months old, and the second member of the control group. C_2's number of correct arithmetic responses decreased from $Baseline_1$ during the New Math Program phase. This subject's behavior is graphically presented in Figure 8.

Student-Subject C_3. C_3 is a fourteen years, seven months old female and a member of the control group. Number of correct arithmetic responses increased slightly during the New Math Program phase. In figure 9, data on the behaviors of C_3 is presented.

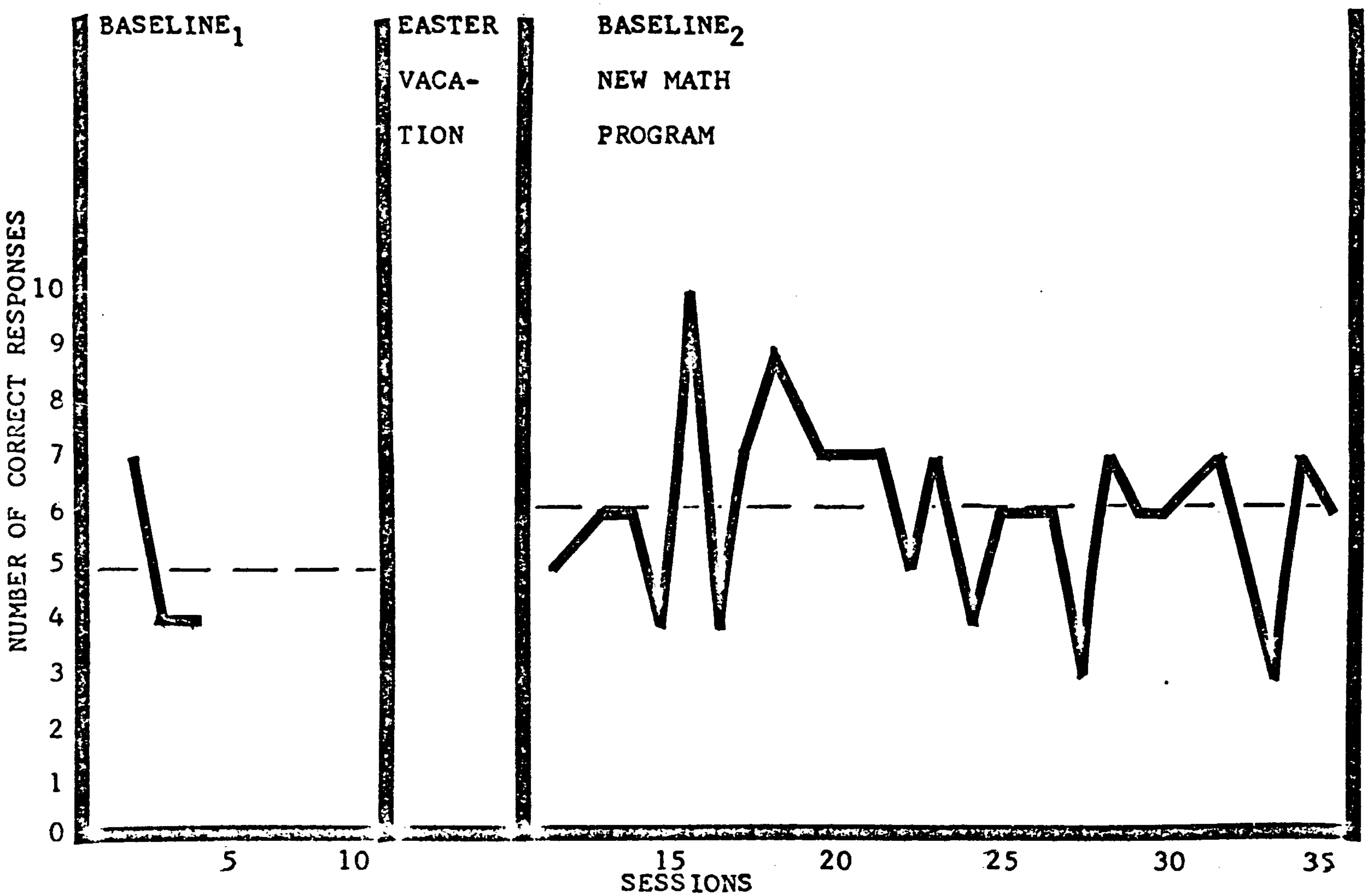

Fig. 7 Student-Subject C_1 : Number of Correct Arithmetic Responses

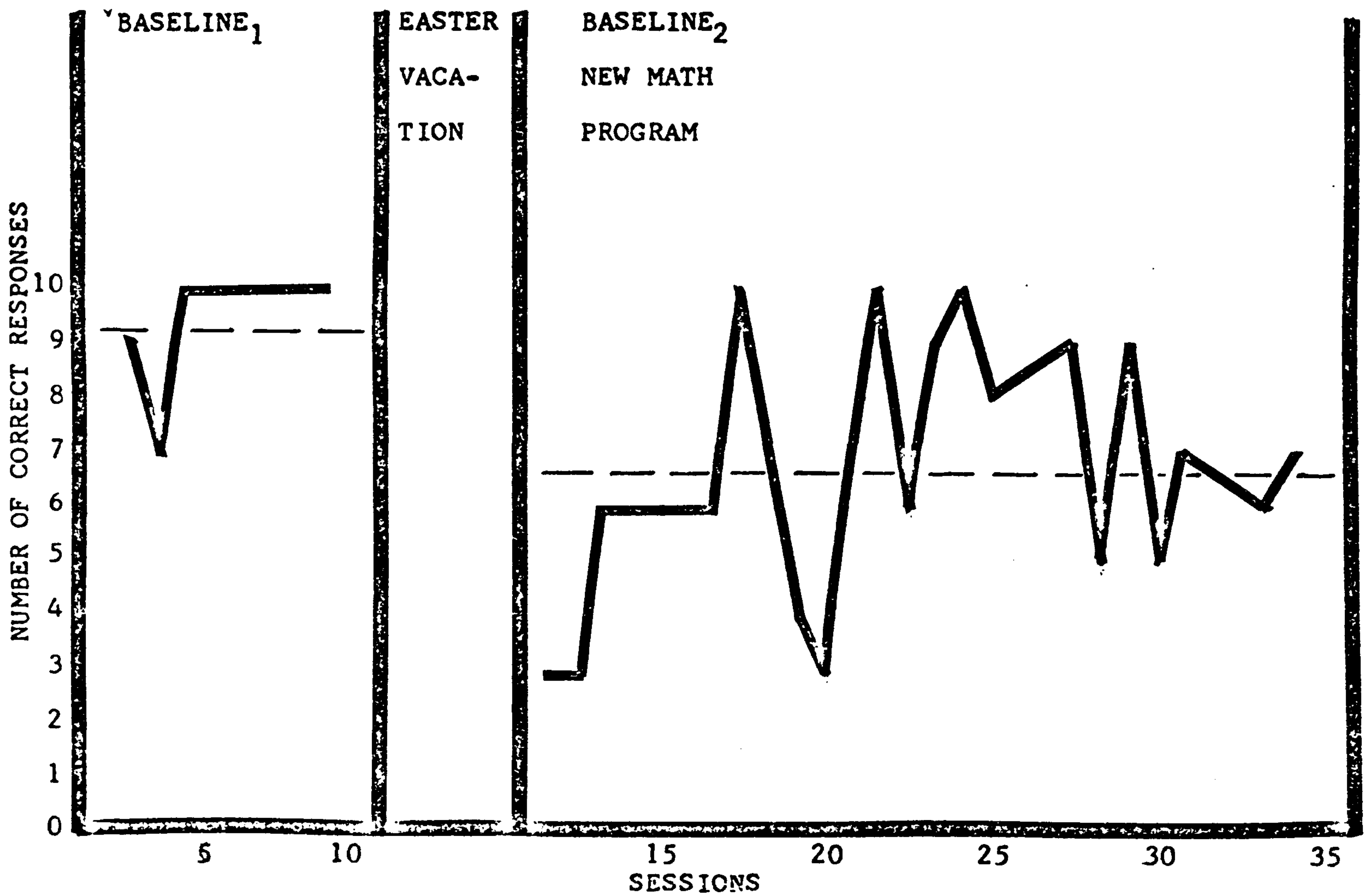

Fig. 8 Student-Subject C_2 : Number of Correct Arithmetic Responses

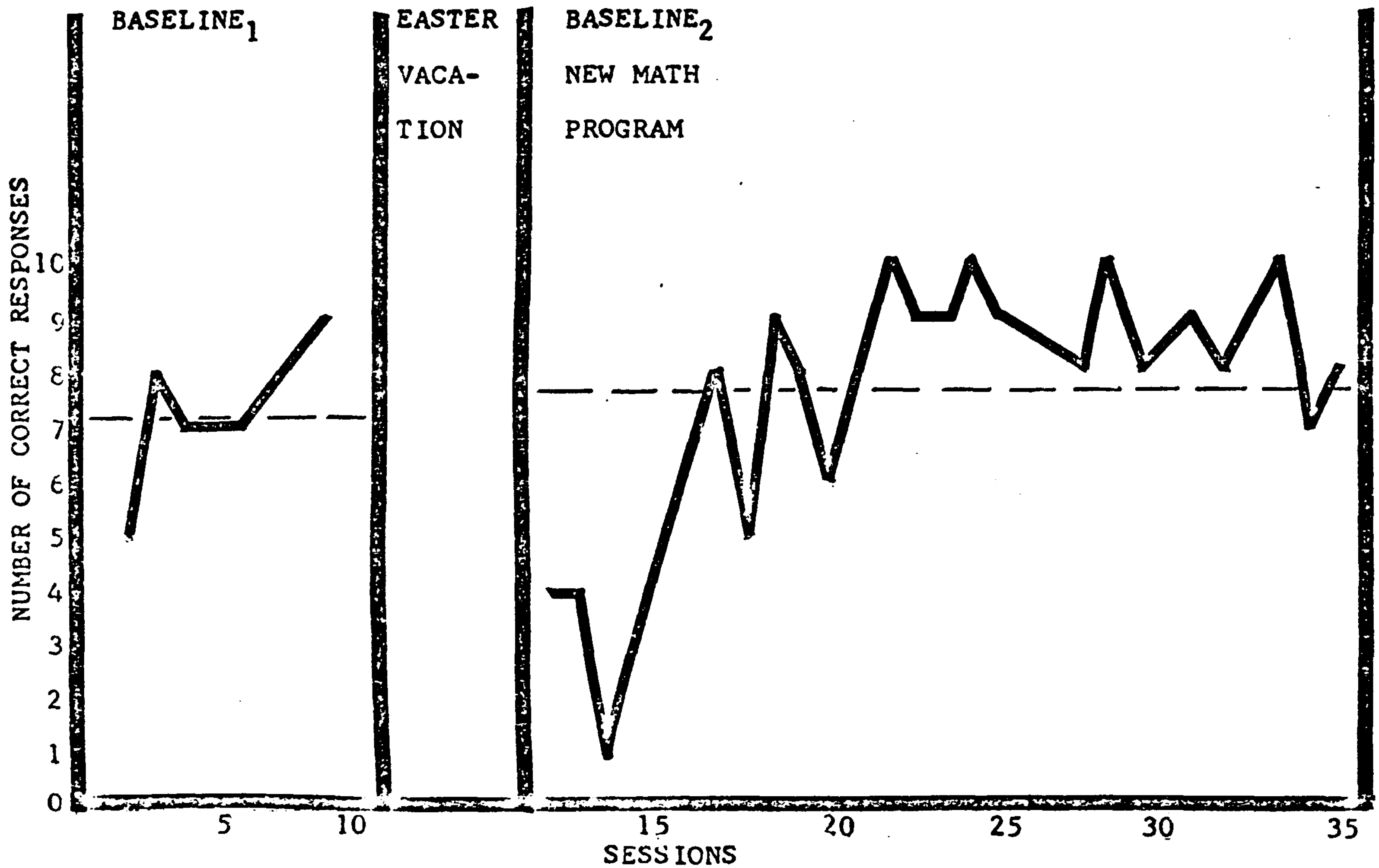

Fig. 9 Student-Subject C_3 : Number of Correct Arithmetic Responses

Teacher A. Teacher A is a member of the experimental group. The New Math Program and the experimental conditions were introduced to Teacher A at the same time they were introduced to student A_1. The amount of time Teacher A spent in arithmetic lessons increased and was maintained from $Baseline_1$ through the New Math Program and Experimental Conditions phases. Figure 10 is a graph of Teacher A's behavior.

Teacher B. Teacher B teaches the replication group of primary educable mentally retarded students. Teacher B decreased the amount of time spent in arithmetic lessons from $Baseline_1$ to the New Math Program. Amount of time spent in arithmetic lessons increased again during the two experimental conditions but did not increase to the amount of time held in $Baseline_1$. Teacher B's behaviors are graphed in Figure 11.

Teacher C. Teacher C teaches the control group of junior high school students. The amount of time that Teacher C spent in arithmetic lessons increased from $Baseline_1$ during the New Math Program phase. The control group did not receive the experimental conditions. Teacher C's behavior is graphed in Figure 12.

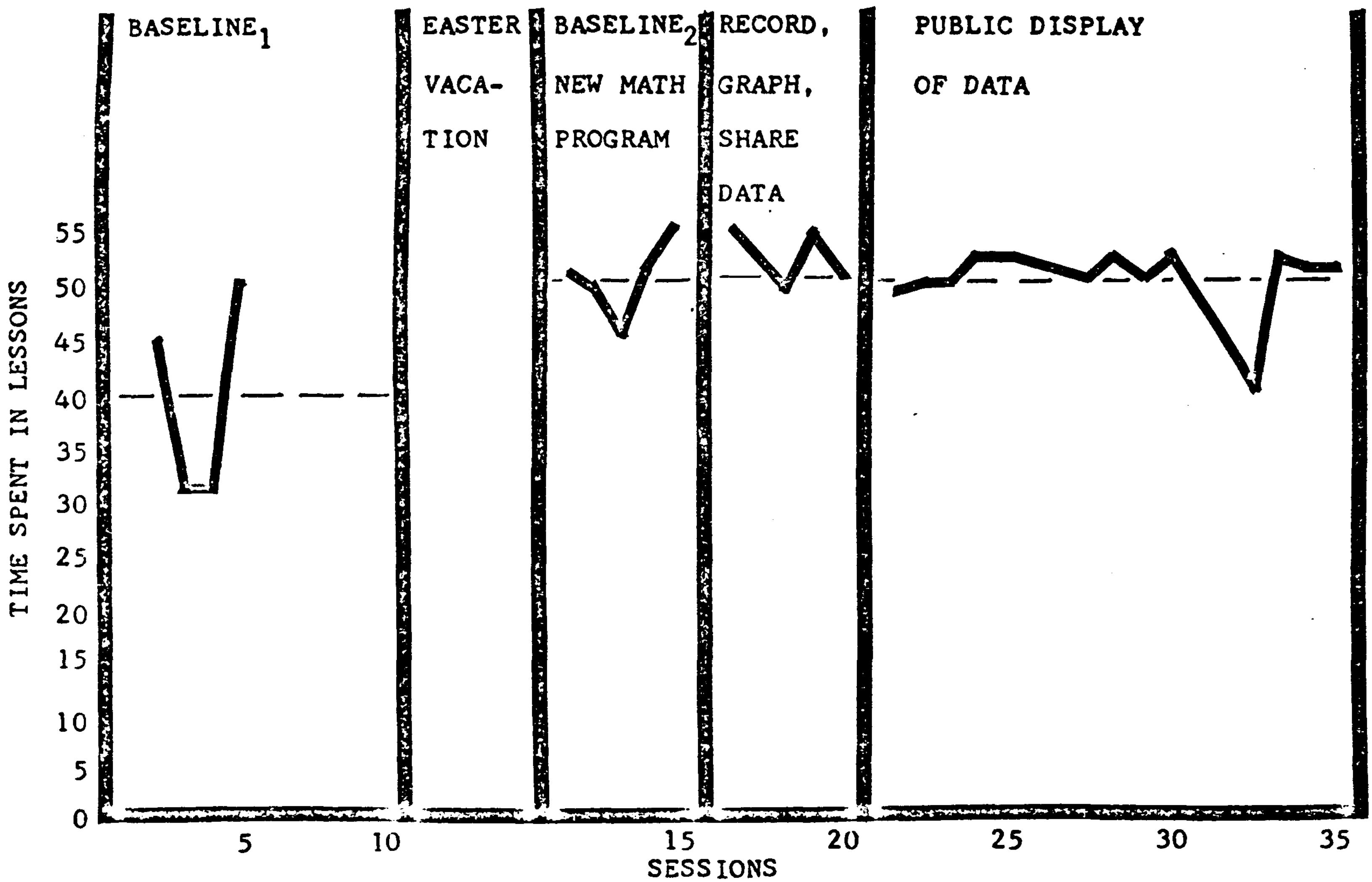

Fig. 10 Teacher A : Amount of Time Spent in Arithmetic Lessons

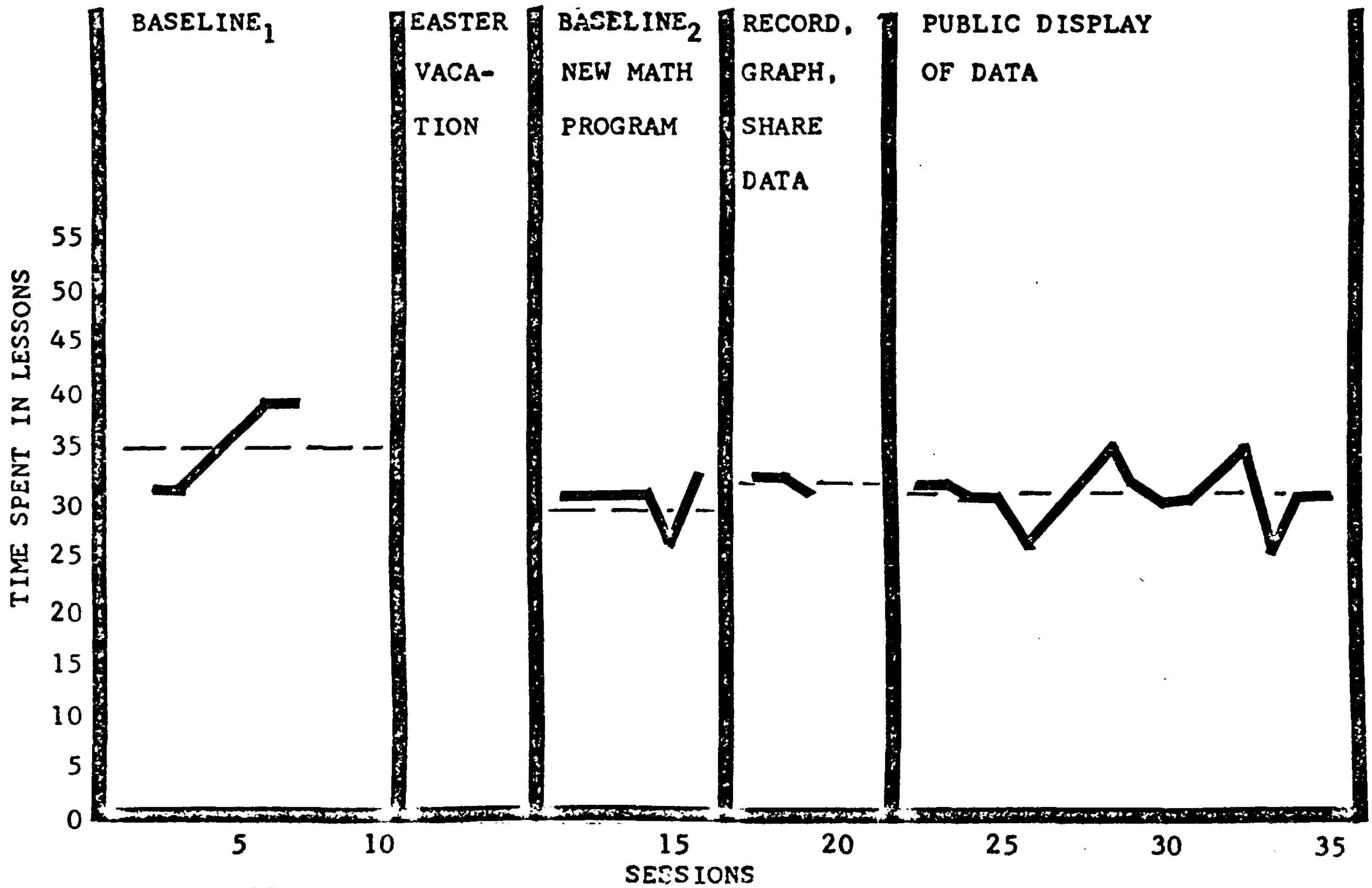

Fig. 11 Teacher B : Amount of Time Spent in Arithmetic Lessons

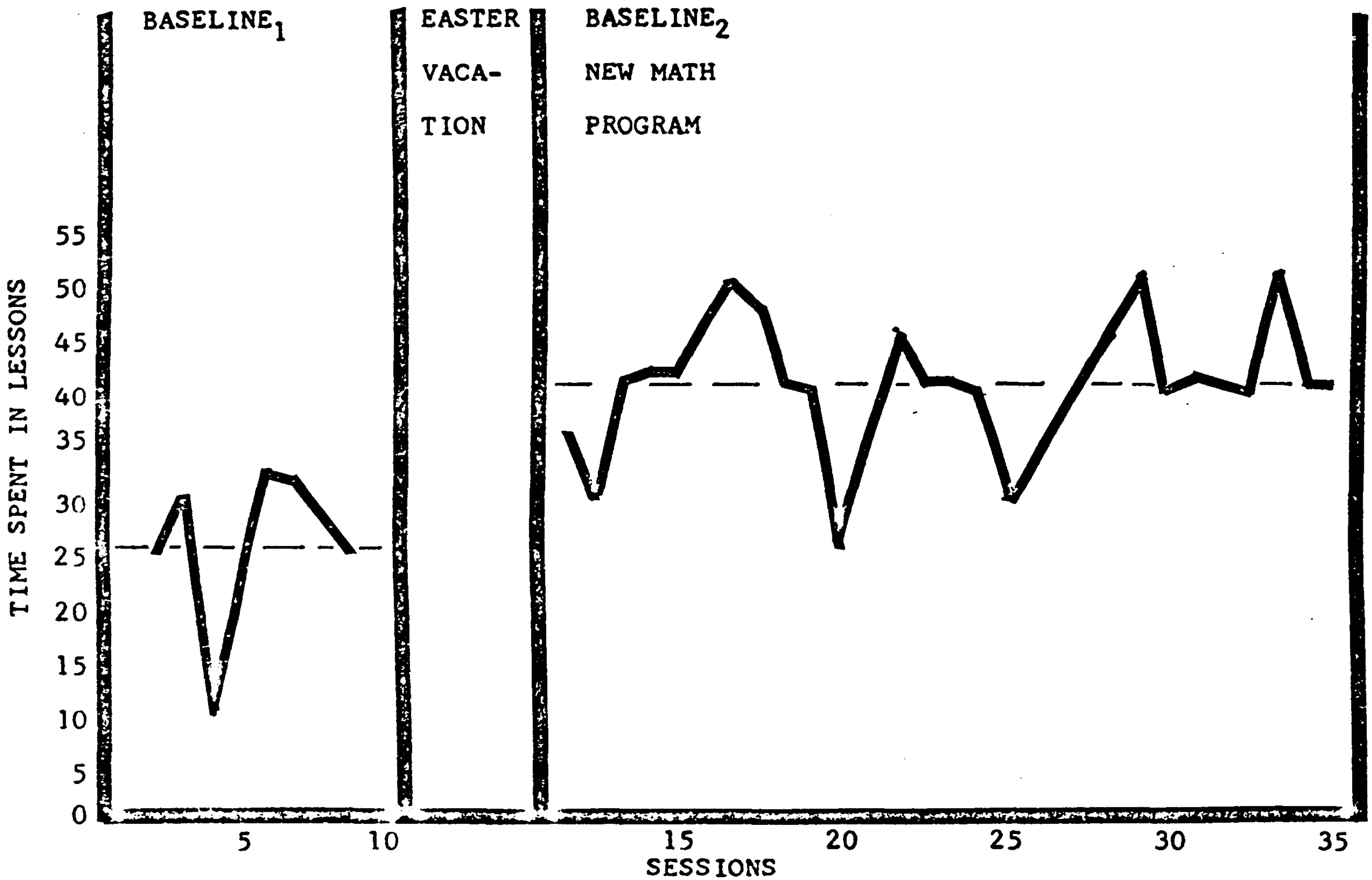

Fig. 12 Teacher C : Amount of Time Spent in Arithmetic Lessons

As the graphs indicate, all the students who received the experimental conditions (Group A and Group B) eventually increased their mean numbers of correct responses over the $Baseline_2$ scores. Originally, the research design did not include $Baseline_2$ which was recorded as the New Math Program phase was introduced. The control group had been incorporated to measure the effects of the New Math Program alone. However, because of the unstable baseline rates of students A_2 and A_3 and the period of absence during Easter vacation, the decision was made to record another baseline in conjunction with the New Math program. Stable rates were established before the experimental conditions were introduced.

Students B_1 and B_2 started with relatively high mean scores (9.66 and 7.57 respectively) during $Baseline_1$. Their low achievement in arithmetic was based on the level at which they were working and not upon the number of errors they were making. When the New Math Program phase was introduced, subtraction of numbers up to ten was combined with the addition facts up to ten. Level of difficulty was increased at this point for these students. This could explain the lower number of correct responses which Students B_1 and B_2 emitted during the New Math Program phase. However, the introduction of the two experimental phases increased the number of correct responses for these students.

The graphed responses indicate that the behaviors of Students A_1, A_2, A_3, and B_1 and B_2 were increased as the Recording, Graphing and Sharing Data with Students was invoked. Student B_3's behavior was very slightly altered by this phase, however B_3's number of correct arithmetic responses increased as the Public Display on a Bulletin Board phase was included in the study.

The public display of performance was only slightly more or less effective for students A_3, B_1 and B_2 than the Recording, Graphing and Sharing of Data alone. It should be recalled that both B_1 and B_2 had high correct response rates, (10.00 and 8.50, respectively) during the previous phase.

Teachers A and C increased the amount of time they spent in arithmetic from Baseline_1 when the New Math Program phase began. This was maintained through the experimental conditions for Teacher A. It was also maintained throughout the remainder of the study for Teacher C, who did not receive the experimental conditions. The amount of time Teacher B spent in arithmetic lessons decreased from Baseline_1.

While a statement of casuality cannot be made when comparing an individual's behavior to him/herself, any maintained increase in the number of correct arithmetic responses emitted by educable mentally retarded students could be interpreted as an indication of successful change. Based on this premise, each student involved in the experimental and replication groups experienced successful changes in behavior when experimental conditions were introduced. Similarly, the assumption was made that an increase in the amount of time teachers spent in arithmetic lessons indicated an increased effort by the teacher to improve student performance. Data on Teachers A and C indicate such an increase in effort.

Subjects Compared Within Groups

Results

The behavior of four groups--experimental Group A, Replication Group B, and control Group C, will be graphically presented in this section; the data of the teachers are also presented as a group. Student Groups A and B represent a multiple baseline design across three subjects. Baseline data were gathered for one behavior across the three persons. Baseline data were again gathered across individuals when the New Math Program was introduced simultaneously to the three subjects. After behavior stabilized (see Operational Definitions, p. 10) across subjects, the first experimental condition was invoked for one subject while baseline conditions were continued for the other subjects. The experimental condition was extended to include the separate individuals. After behavior stabilized for the first subject in the first experimental condition, the second experimental condition was invoked. This second experimental condition also was extended to include the separate individuals. Figure 13 is a graph of the multiple baseline data of experimental Group A. Figure 14 is a graph of the data of replication Group B.

Figure 15 is a graph combining the recorded behaviors of the teachers. Data indicates the amount of time the teachers spent in arithmetic lessons. Teachers A and C increased the mean amount of time they spent in arithmetic lessons from $Baseline_1$. The increase began when the New Math Program was introduced. Teacher B decreased the mean amount of time spent in arithmetic lessons from $Baseline_1$.

Figure 16 is a graph of the behaviors of the control group. The Wilcoxon Signed Rank Test (Hollander and Wolfe, 1973, p. 27) was applied separately to the data of the teachers and to the student-subjects in the control group. This test was used to measure the effects, if any, of the New Math Program on the behaviors of the subjects.

Subjects Compared Across Groups

Results

Data which were gathered on student-subjects who received the treatment conditions are graphically presented in Figure 17. These are the students in Groups A and B (experimental and replication groups, respectively). $Baseline_1$ is not included in this section since statistical treatment was applied only to data from the New Math

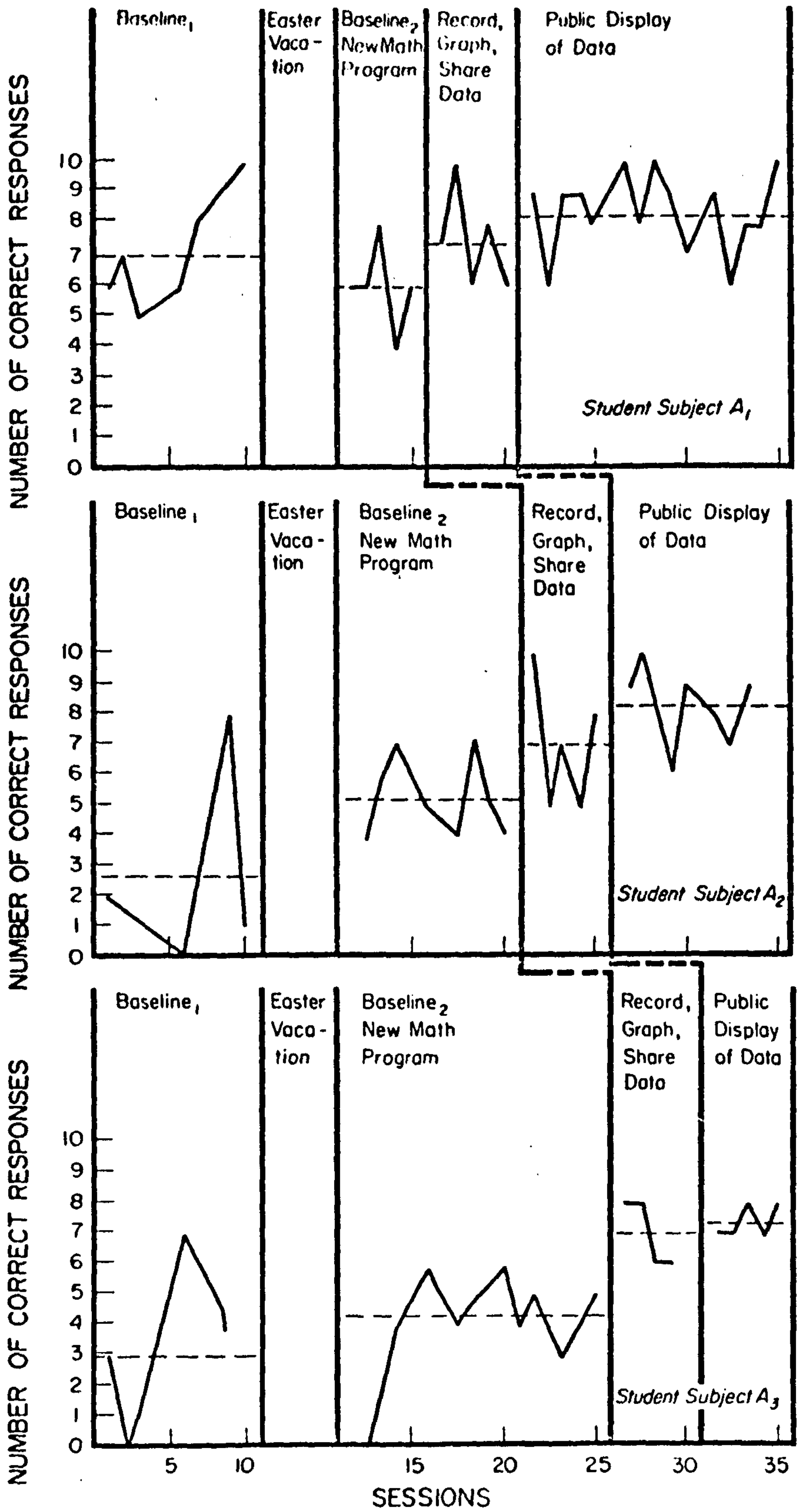

Fig. 13 Multiple Baseline Data : Group A

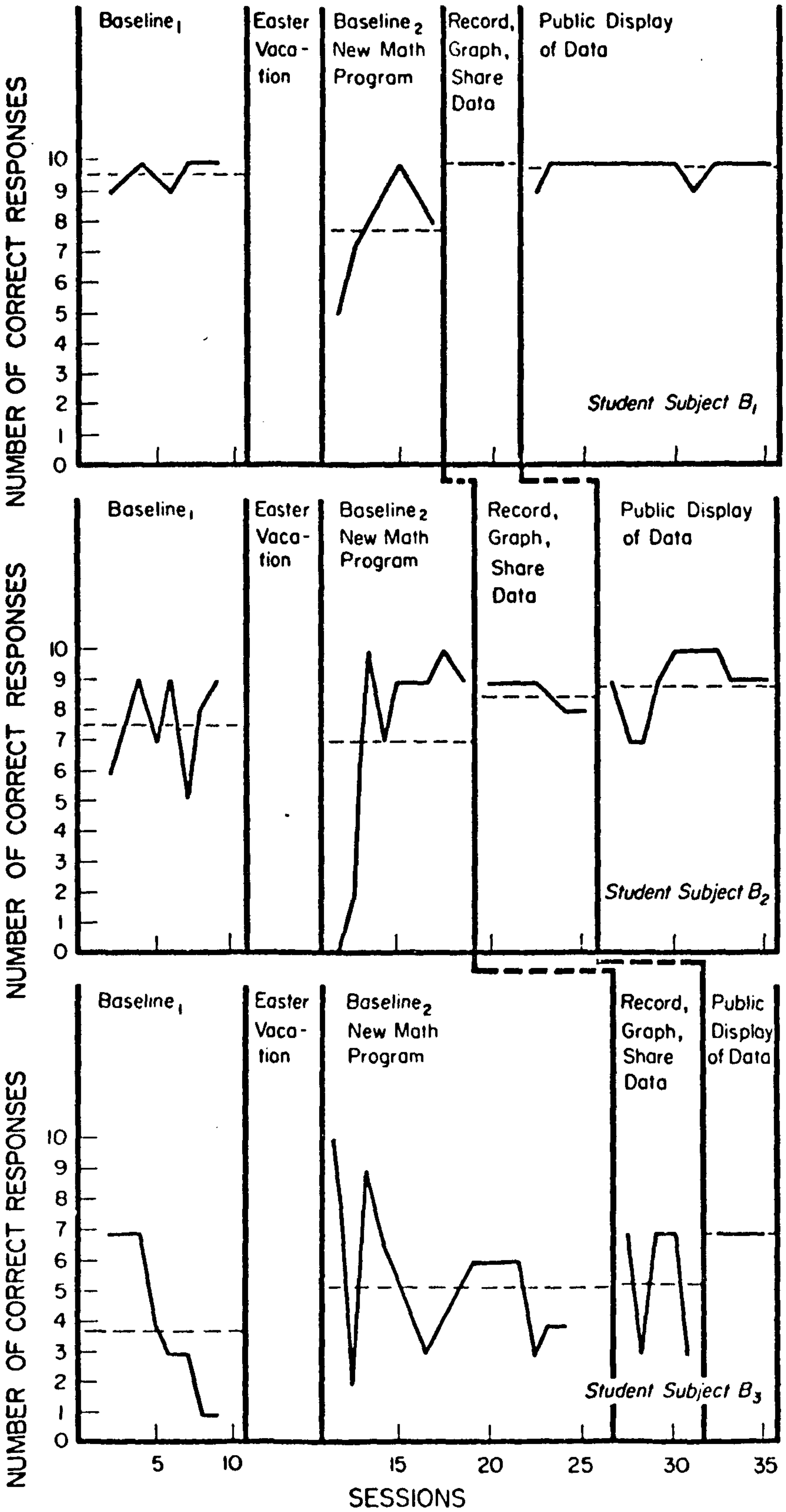

Fig. 14 Multiple Baseline Data : Group B

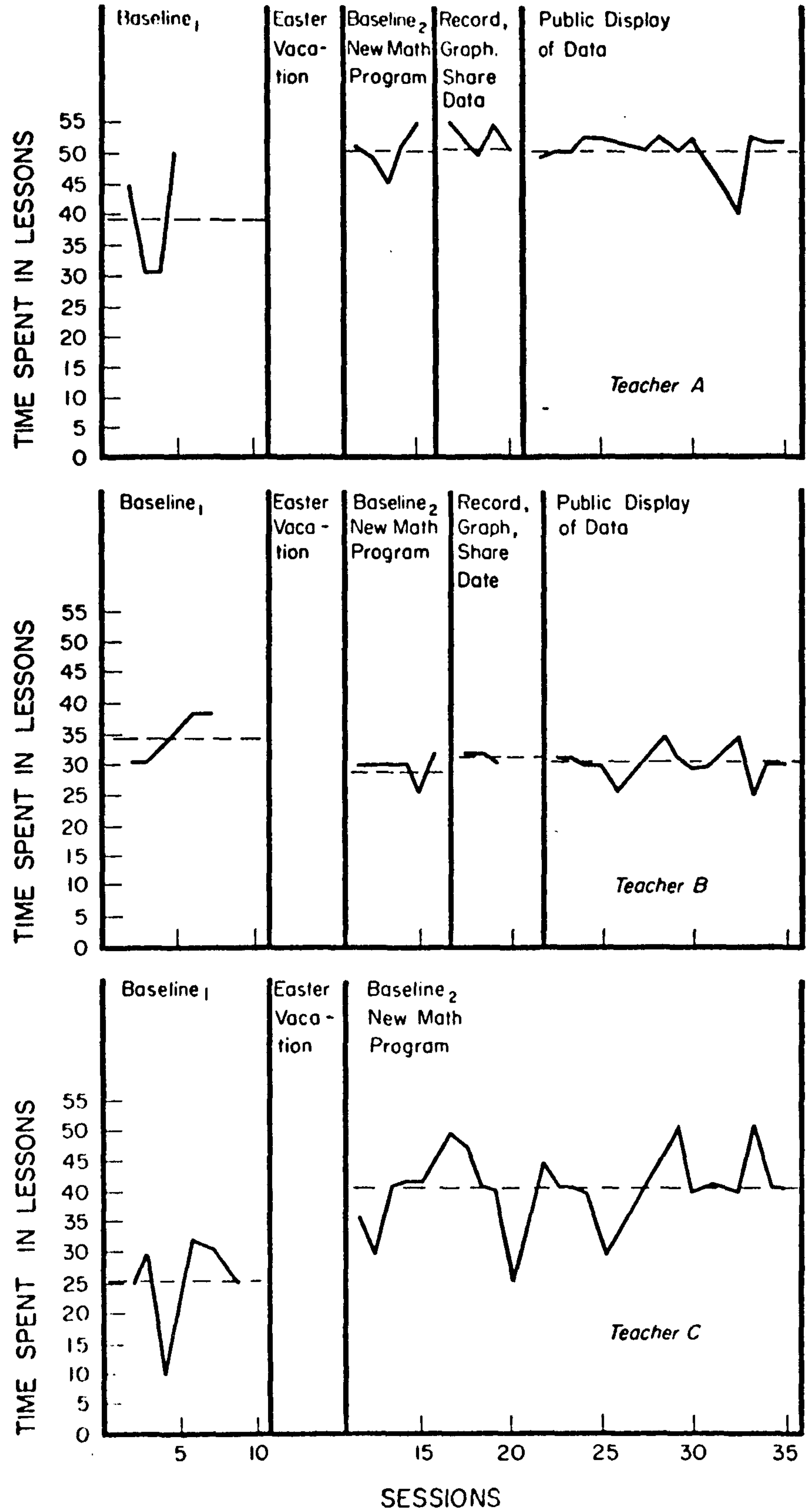

Fig. 15 Combined Recorded Behaviors of Teachers

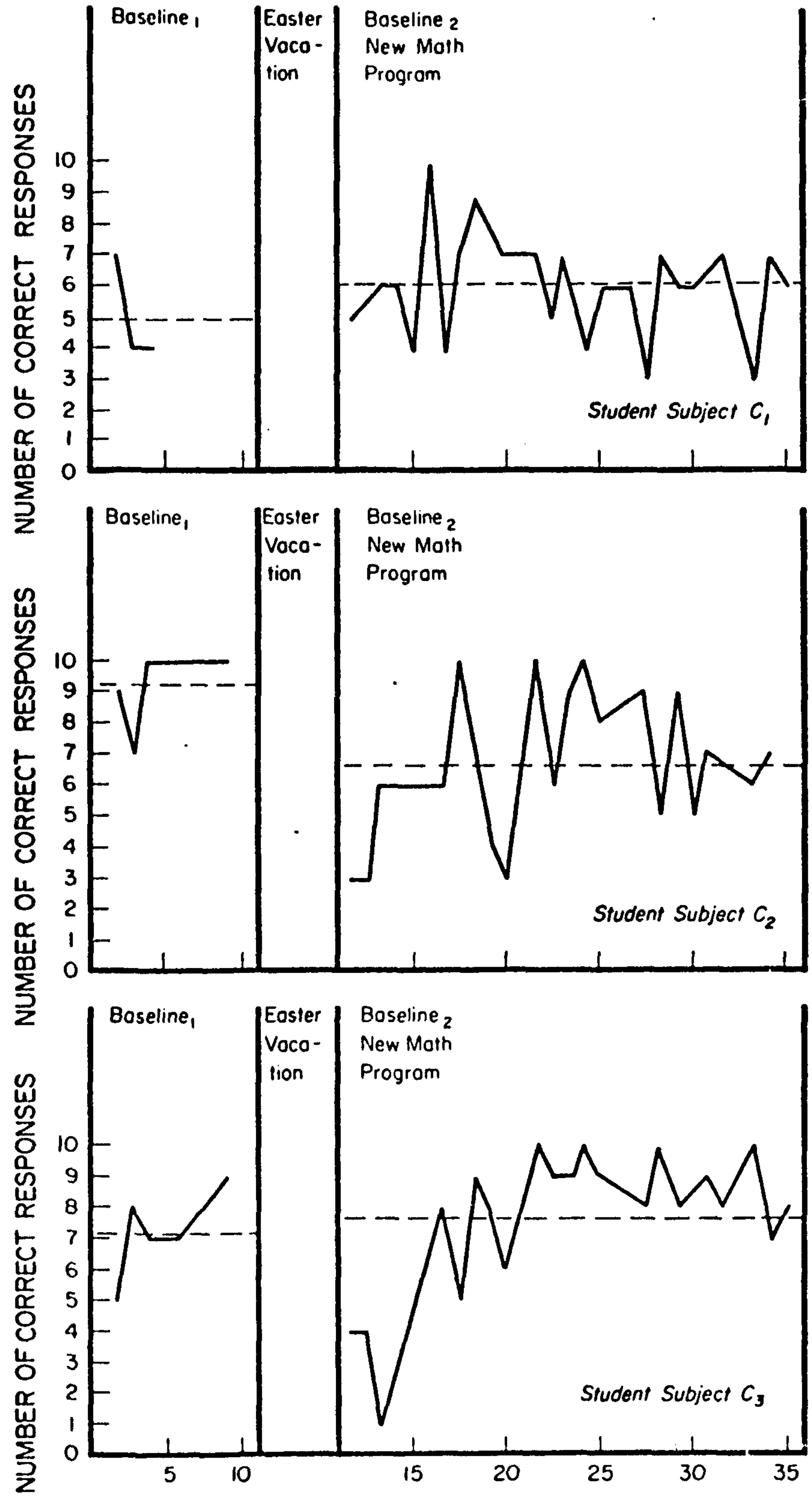

Fig. 16 Combined Recorded Behaviors of the Control Group : Group C

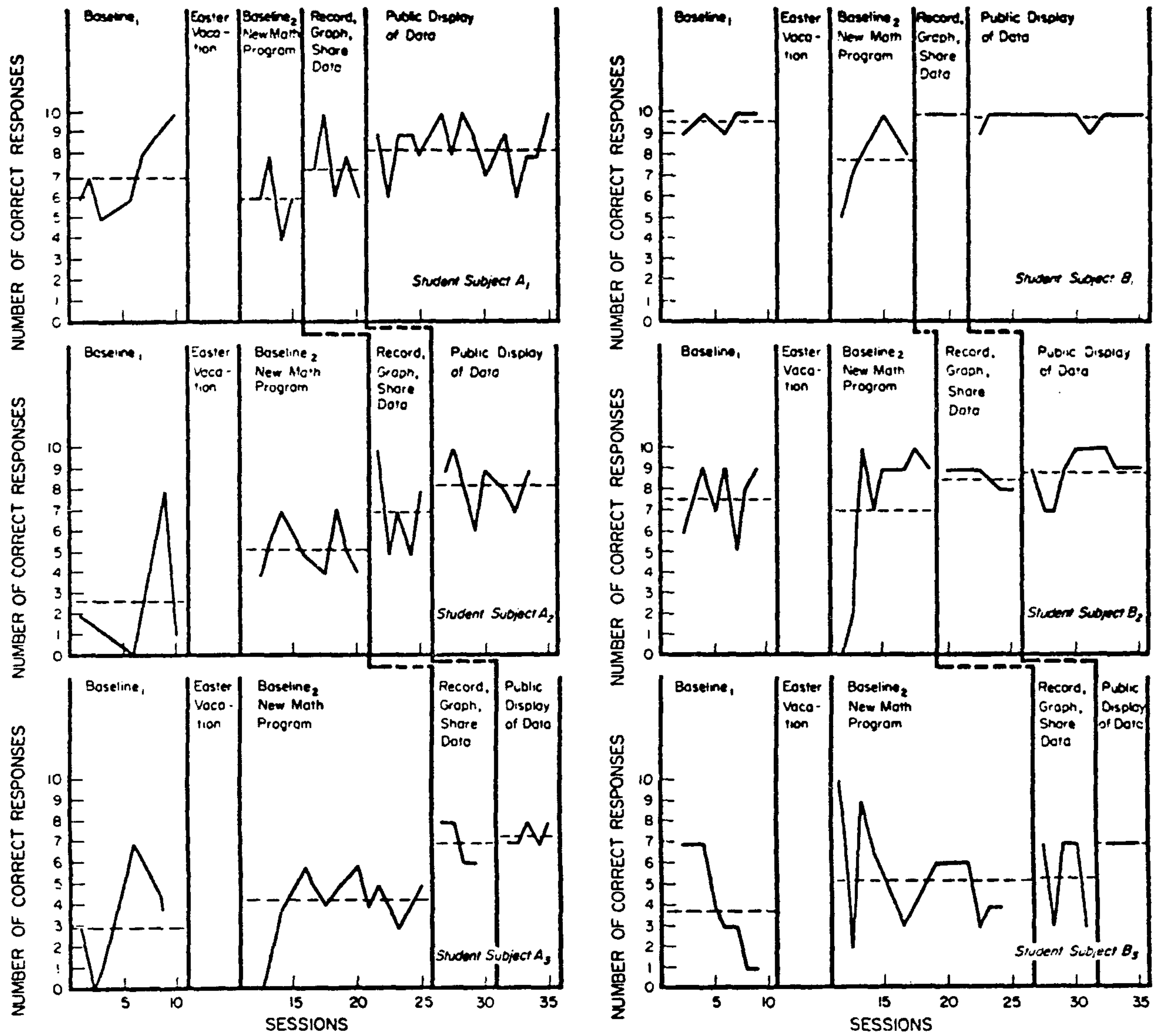

Fig. 17 Subjects Compared Across Groups : Groups A and B

Program phase and the two treatment phases. As stated previously in this chapter, all the students in Groups A and B increased their mean numbers of correct responses over $Baseline_2$. All six students increased mean numbers of correct responses when the first experimental condition was introduced. During the second experimental condition, five of the six students further increased mean numbers of correct responses. The sixth student slightly decreased his mean number of correct responses during the second experimental phase.

The Friedman Test was applied to the data of Groups A and B. When the null hypothesis was not found tenable, a multiple comparison procedure was applied to determine which treatments differed significantly (Hollander and Wolfe, 1973, p. 151).

Hypotheses

The hypotheses of the study are restated here together with statements as to their tenability in light of statistical results. A full discussion will follow in the next chapter of this study.

Ho_1: There will be no statistically significant change ($\alpha = .05$) in the number of correct arithmetic responses emitted by three low achieving students when knowledge of daily quiz scores are recorded and graphed and shared with the student by the teacher. (Accepted)

Ho_2: There will be no statistically significant change ($\alpha = .05$) in the number of correct arithmetic responses emitted by three low achieving students when the recorded and graphed results of number of correct responses on daily arithmetic quizzes are publicly displayed on the classroom bulletin board. (Rejected) $\underline{\alpha}$ = .009

Ho_3: There will be no statistically significant change (α = .05) in the amount of time the teacher spends in arithmetic lessons when she/he records, graphs, and shares the number of correct responses on daily arithmetic quizzes with three low achieving students. (Accepted)

Ho_4: There will be no statistically significant change (α = .05) in the amount of time the teacher spends in arithmetic lessons when the number of correct responses on daily arithmetic quizzes of three low achieving students are recorded, graphed and publicly displayed on the classroom bulletin board. (Accepted)

CHAPTER V

SUMMARY, DISCUSSION AND IMPLICATIONS

Summary

Accountability has become a major concern of educators and legislators within the last ten years. Recent developments that appear to have influenced the current emphasis and concern with accountability are: the increased proportion of income that is taxed for school support; the recognition that many youth are failing to meet standards of literacy demanded for employment (Tyler, 1971, p. 1); demands for equal educational opportunities by parents of handicapped children and parents of children in minority groups (Vergason, 1973); the dissatisfaction of some special educators with the labeling, and subsequent education of mildly retarded students (Dunn, 1968); and general increased parental concern with what children are learning and why they are being taught whatever they are being taught (Milliken, 1971, p. 18).

The big issue moves from the recognition that educators must meet the obligation of accountability to determining the best procedure for meeting that obligation (Lessinger, 1971, pp. 28-41; Milliken, 1971, pp. 18-20; Peterson, 1971, pp. 20-24; Nyquist, 1971, pp. 24-27). The purpose of this study was to suggest one approach to solving this problem; namely, by applying a scientific behavior analysis technique (i. e., recording and graphing behavior and sharing results with students) within classrooms for students classified as educable mentally retarded; and further, to measure the effects of the technique on the performance of both the students and their teachers.

Educable mentally retarded students and their teachers were selected as subjects for two reasons: (1) because special education was one of the first areas in which the need for accountability was recognized (Dunn, 1968; Vergason, 1971, pp. 28-42); and (2) because research results suggest that knowledge of performance, which is a part of the particular behavior analysis technique used in this study, may be a reinforcer where uncertainty is high, or where confidence is low (Geis and Chapman, 1972). A behavior analysis technique was selected as the approach for critically analyzing the teaching-learning process because it is easily employed by classroom teachers. One who reads the literature finds that many writers (Lessinger, 1972, pp. 7-15; Nyquist, 1971, pp. 24-27; Tyler, 1971, pp. 1-6; Porter, 1971, pp. 42-51; Reddick, 1972; Vergason, 1973) suggest that the responsibility of the accountability process rests with the teacher.

In this investigation an attempt was made to determine whether knowing how a student was progressing would alter the behaviors of both the student and his/her teacher. To provide data for this purpose, the behaviors of three EMR teachers, and three students which each teacher selected, were recorded during baseline and two experimental treatment phases. Subjects were designated the experimental group (Group A), the replication group (Group B) and the control group (Group C). Each group consisted of one EMR teacher and three EMR students who were described as low achievers in arithmetic (see Operational Definitions, p. 10). Comparisons were made within individual behaviors, within group behaviors, and across group behaviors.

The findings of this study were presented narratively and graphically. Comparisons were made of the mean numbers of correct arithmetic responses of each student during baseline and treatment phases. It was hypothesized that there would be no statistically significant ($\alpha = .05$) change in the number of correct arithmetic responses emitted by three low achieving students when knowledge of daily quiz scores were recorded and graphed and shared with the student by the teacher. The data from the experimental (Group A) and replication (Group B) groups support the null hypothesis.

A second hypothesis was that there would be no statistically significant ($\alpha = .05$) change in the number of correct arithmetic responses emitted by three low achieving students when the recorded and graphed results of number of correct responses on daily arithmetic quizzes were publicly displayed on the classroom bulletin board. This hypothesis was rejected for Group A ($\underline{\alpha}$ = .028), but not for Group B, when data from each group were treated independently. However, when data from the two groups were combined, the null hypothesis was rejected ($\underline{\alpha}$ = .042).

Comparisons of students' behaviors within individuals indicated that all the students who received the experimental conditions (Group A and Group B) eventually increased their mean numbers of correct responses over Baseline_2.

Comparisons were also made of the amount of time each teacher spent in arithmetic lessons during each phase of the study. A third hypothesis stated that there would be no statistically significant change in the amount of time the teacher spent in arithmetic lessons when she/he records, graphs, and shares the number of correct responses on daily arithmetic quizzes with three low achieving students. This hypothesis was accepted.

A fourth hypothesis was that there would be no statistically significant change in the amount of time the teacher spends in arithmetic lessons when the number of correct responses on daily arithmetic quizzes of three low achieving students are recorded, graphed and publicly displayed on the classroom bulletin board. This hypothesis was also accepted. However, the comparison of individual teachers to themselves (see figures 10, 11 and 12) suggests that Teacher A and Teacher C increased the amount of time they spent in arithmetic lessons when the New Math Program phase was introduced and that Teacher B decreased the amount of time spent in arithmetic lessons when the New Math Program was introduced.

Discussion

The results of this study suggest that the public display of student performance on classroom bulletin boards does significantly change the arithmetic performance of educable mentally retarded students. Statistical treatment indicated that the second, third, and fourth null hypotheses should be accepted. However, comparisons of student and teacher behavior, both within individuals and within groups, indicate that while some treatment effects may not be statistically significant, they may nevertheless be important.

In the multiple baseline design used with Group A and Group B, the response frequency is expected to change as the experimental condition is extended to include separate individuals. This is intended to demonstrate that the behavior of each subject does not change until he/she is included in the experimental condition. The effects of the experimental operations may be immediately observed on response rates (Kazdin, 1973).

With the exception of Student B_3, the data collected on the behaviors of the Group A and Group B participants indicates that the behavior of each subject did not change until the first experimental condition was invoked. Response rates again increased for students A_1, A_2, A_3, B_2 and B_3 when the second experimental condition was introduced. Response frequency changed, i. e. increased above the New Math Program baseline ($Baseline_2$) for all the students either during the first or second experimental phase of the study, or during both phases. This indicates that the experimental conditions were altering the behavior of the subjects.

In the comparison of individual student-subjects to themselves, data indicates that the number of correct arithmetic responses which students made increased during either one, or both, of the experimental treatments. In the multiple baseline design used in this study, a statement of causality cannot be made when looking at differences in an individual's performance, but an indication of successful change can be given. If it is desirable to increase the number of correct arithmetic responses made by educable mentally retarded students, then each student in this study experienced a successful change in behavior.

The results of the study, when data is statistically treated, also suggest that the treatment conditions had no effects on teacher behavior. However, a visual perusal of teacher data does indicate changes in behavior when the New Math Program was introduced. Teachers A and C increased, and maintained the increase in, the amount of time they spent in arithmetic from $Baseline_1$ when the New Math Program phase began. The knowledge that another person was interested in their students' progress may have served as a motivation to these teachers. The interest of other people in the progress of students is inherent in the public demand for accountability. Perhaps the teachers' recognition of this demand would have the same effect that the New Math Program had on teacher behavior.

Because of special interest in the serendipity discovered in teacher behavior, the Wilcoxon Signed Rank Test was applied to the teachers' data. The effect of the New Math Program was found to be statistically significant at the .250 level.

The Wilcoxon Signed Rank Test was also applied to the data of the student-subjects in the control group to determine the effects, if any, of the New Math Program alone on student behavior. The effects of the New Math Program alone were significant at the .625 level; meaning that 62 1/2 times out of 100, this effect could be produced by chance. The high probability of this effect being produced by chance would negate any suggestion that the New Math Program was affecting the results attributed to the experimental conditions which Groups A and B experienced.

Implications

The results of this study suggest that teachers can be trained to use a behavior analysis technique which provides a visual representation of pupil progress. Such visual representations can be easily interpreted by students, teachers, parents, and the community which financially supports the school system. Results further suggest that this visual representation of pupil progress changes the behavior of the students whose performances are expressed by it. This is especially useful information because the change in performance was in the direction of increasing a socially desirable behavior. The desirable behavior increased for each student irrespective of any individual differences there may have been in such characteristics as age, sex, frustration levels, and type of home environment. No attempt was made to match students in these characteristics.

A second implication based on the results of this study is that teachers will spend more time in instruction in a particular area when an "outside agent" is involved in co-planning individual programs in that area and in discussing pupil progress. The results of this study may suggest that external auditing agencies may cause changes in teacher behavior.

One important implication is that successful changes in pupil progress need not be, and perhaps should not be, based on the statistical significance of the effect of a particular treatment condition. Had statistical significance been the only criterion for determining the effects of treatments in this study, much important information on pupil progress and teacher behavior would have been lost. Accountability, that is, the critical analysis of the teaching-learning process, should result in the assessment of the performance patterns of each student; that is, his or her strengths and weaknesses in every area of learning. Inherent in the critical analysis of the teaching-learning process is the development of a system for correcting the educational deficiencies which have been revealed (Milliken, 1971, p. 18). Data should be compared within individuals, even though that data cannot be treated statistically.

Another implication of this study is that the behavior analysis technique and the research designs used in this study may be fundamental to meeting the obligation of accountability in education. They can be used to measure individual behavior and to support statements of causality. They are easily applied by teachers and result in visual representations of data which can be readily interpreted by an unsophisticated audience.

Suggestions for future research would be: (1) to use the same treatment conditions, or others which simulate making pupil progress known to the public, with students other than retardates; and (2) to study the effects of the public display of pupil performance on teacher behavior when teachers are aware that their instructional behaviors are also being observed. Other possible studies could investigate the effects of knowledge of student progress on the community which receives regular public reports. One desired reaction, of course, is continued or increased financial support. This could be one measure of public support. Other studies could measure public support of school activities, school authorities, and/or school rules.

Since the results of this study suggest that knowledge of student performance can cause changes in pupil behavior under certain conditions, more studies should be conducted to determine if there are conditions which might cause undesirable changes in behavior. Before the accountability process terminates in revealing pupil progress at public meetings or in the local newspaper, research should be conducted to determine if these procedures could have aversive affects on student and/or teacher behavior.

BIBLIOGRAPHY

Allen, E.; Hart, B.; Buell, J.; Harris, F.; and Wolf, M. "Effects of Social Reinforcement on Operant Crying." Journal of Exceptional Child Psychology, 1964, 1, pp. 145-153.

Anderson, H. M. "A Study of Certain Criteria of Teacher Effectiveness." Journal of Experimental Education, 1954, 23, pp. 41-71.

Anderson, R. C.; Kulhavy, R. W.; and Andre, T. "Feedback Procedures in Programmed Instruction." Journal of Educational Psychology, 1970.

Baer, D.; Wolf, M.; and Risley T. "Some Current Dimension of Applied Behavior Analysis." Journal of Applied Behavior Analysis, 1968, 1, pp. 91-97.

Bapst, Jennis J. "The Effect of Systematic Student Response upon Teaching Behavior." The Biology Teacher, July, 1972.

Bain, Helen. "Self-Accountability Can Make the Difference." Teacher Educator, v 7, n 2, Winter 71-72.

Barr, A. S. Journal of Experimental Education, June, 1948, (Journal of Educational Research, May, 1952, Vol XLV, No. 9).

Bijou, S. W. "Experimental Studies of Child Behaviors, Normal and Deviant," in Krasner, L. and Ullman, L. P. (Eds.) Research in Behavior Modification, New York: Holt, Rinehart, and Winston, 1966.

Birch, D. R. "Guided Self-Analysis and Teacher Education." Unpublished doctoral dissertation, University of California at Berkeley, 1969.

Broden, M.; Hall, V. R.; and Mitts, B. "The Effects of Self-Recording on the Classroom Behavior of Two Eighth-Grade Students." Journal of Applied Behavior Analysis, Winter, 1971.

Brookover, W. B. "The Relation of Social Factors to Teaching Ability." Journal of Experimental Education, 1945, 13, pp. 191-205.

Brown, Louis F. "Evaluation: Some Informal Techniques for Classroom Use." Strategies for Teaching Exceptional Children, Love Publishing Co., Denver, Colorado, 1972.

Campbell, D. T., and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Rand McNally and Company, Chicago, Ill., 1963.

Campeau, P. L. "Test Anxiety and Feedback in Programmed Instructions." Journal of Educational Psychology, 1968, 59 (3), pp. 159-163.

Clark, Hewitt B. "A Program of Delayed Consequences for the Management of Class Attendance and Disruptive Classroom Behavior on 124 Special Education Children." Behavior Analysis and Education, 1972, The University of Kansas, 1972.

Cooper, John O. Handbook of Measurement and Analysis for Directive Teachers, 1974.

Cossairt, Ace, Hall, R. Vance; and Hopkins, V. L. "The Effects of Experimenter's Instructions, Feedback, and Praise on Teacher Praise and Student Attending Behavior." Journal of Applied Behavior Analysis, Spring, 1973.

Cummings, L. L.; Schwab, Donald; and Rosen, Marie. "Performance and Knowledge of Results as Determinants of Goal Setting." Journal of Applied Behavior Analysis, v 55, n 6, pp. 526-530, December, 1971.

Domas and Tiedman. Journal of Experimental Education, June, 1950 (entire issue).

Dunn, Lloyd. "Special Education for the Mildly Retarded--Is Much of It Justifiable?" Exceptional Children, Vol. 35, September 1968.

Feldhusen, J., and Birt, A. "A Study of Nine Methods of Presentation of Programmed Learning Material." Journal of Educational Research, 1962, 55(9), pp. 461-466.

Ferguson, George A. Statistical Analysis in Psychology and Education, McGraw-Hill Book Co., New York, N. Y., 1964.

Ferster, C. B., and Skinner, B. F. Schedules of Reinforcement, New York: Appleton-Century-Crofts, Inc., 1957.

Geis, G., and Chapman, R. "Knowledge of Results and Other Possible Reinforcers in Self-Instructional Systems." Motivation and Learning, Educational Technology Publication, Englewood Cliffs, New Jersey, 1971.

Geis, G.; Jacobs, W.; Spencer, D.; and Nielson, S. "The Role of the Printed Answer in Programmed Instruction." NSPI Journal, 1970, 9(6), pp. 8-11.

Glaser, R., and Taber, J. I. Investigation of the Characteristics of Programmed Learning Sequences, University of Pittsburgh, Programmed Learning Lab, Coop., Research Project 691, Chap. 4, 1961.

Goetz, E., and Baer, D. "Descriptive Social Reinforcement of 'Creative' Block Building by Young Children." A New Direction for Education: Behavior Analysis, 1971, University of Kansas, 1972, pp. 223-234.

Goetz, E., and Salmonson, M. "The Effect of General and Descriptive Reinforcement on 'Creativity' in Easel Painting." Behavior Analysis and Education, 1972, The University of Kansas, 1972.

Goyham, R. E. "Personality and Teaching Efficiency." Journal of Experimental Education, 1945, 14, pp. 157-165.

Hall, R. Vance. Managing Behavior I, Behavioral Modification: The Measurement of Behavior, H & H Enterprises, Inc., Lawrence, Kansas, 1971.

Hall, R. Vance. "Responsive Teaching: Focus on Measurement and Research in the Classroom and the Home." Strategies for Teaching Exceptional Children, Love Publishing Co., Denver, Colorado, 1972, pp. 403-415.

Hall, R. Vance. "Training Teachers in Classroom Use of Contingency Management." Educational Technology, April, 1971, pp. 33-38.

Hall, R. Vance, and Broden, M. "Behavior Changes in Brain-Injured Children through Social Reinforcement." Journal of Experimental Child Psychology, 1967, 5, pp. 463-479.

Hall, R. Vance; Lund, D.; and Jackson, D. "Effects of Teacher Attention on Study Behavior." Journal of Applied Behavior Analysis, 1968, 1, pp. 1-12.

Hamblin, R.; Hathaway, C.; and Wodarski, J. "Group Contingencies, Peer Tutoring and Accelerating Academic Achievement." A New Direction for Education: Behavior Analysis, 1971, University of Kansas, 1971, pp. 41-53.

Hellfritsch, A. G. "A Factor Analysis of Teacher Abilities." Journal of Experimental Education, 1945, 14, pp. 166-169.

Hewett, F.; Taylor, F.; and Artuso, A. The Engineered Classroom: An Innovative Approach to Education of Children with Learning Problems, An unpublished report for Warren J. Aaronson, Director, Title III program, Project Centers Branch, Bureau of Education for the Handicapped, United States Office of Education, 1969.

Hillman, B. W. "Two Methods of Facilitating Classroom Learning: Implications for the Counselor Consultant." The School Counselor, 1969, 77, No. 2, pp. 131-137.

Holland, J. "Research on Programming Variables," In R. Glaser (Ed.) Teaching Machines and Programmed Learning, II: Data and Directions, Washington, D.C., DAVI, National Education Assoc., 1965.

Hundal, P. S. "Knowledge of Performance as an Incentive in Repetitive Industrial Work." Journal of Applied Psychology, 53, pp. 224-226.

James, Barbara. Changes in Self-Understanding as a Result of Participating in Group Training, DHEW, Washington, D.C., West Virginia University, Morgantown, W. Va., Nov. 1972.

Jastak, J. F.; Bijou, S. W.; and Jastak, S. R. "Wide Range Achievement Test." Guidance Associates, Wilmington, Delaware, 1965.

Jayne, C. D. "A Study of the Relationship Between Teaching Procedure and Educational Outcomes." Journal of Experimental Education, 1945, 14, pp. 101-134.

Johnston, Maerteus, and Schooley. "When to Correct Arithmetic Problems--A Critical Variable." School Science and Mathematics, 1969, 69, pp. 799-805.

Jones, R. D. "The Prediction of Teaching Efficiency from Objective Measures." Journal of Experimental Education, 1946, 15, pp. 85-99.

Kazdin, A. E. "Methodological Assessment Considerations in Evaluating Reinforcement Programs in Applied Settings." Journal of Applied Behavior Analysis, 1973, 3, pp. 517-539.

Kerlinger, F. N. Foundations of Behavioral Research. Holt, Rinehart, and Winston, Inc., New York, N.Y., 1971.

Knight, M.; Hasazi, S.; and McNeil, M. "A Home Based Program for the Development of Reading Skills for Preschoolers." A New Direction for Education: Behavior Analysis, 1971, University of Kansas, 1971, pp. 223-234.

Krumboltz, J. D., and Weisman, R. G. "The Effect of Intermittent Confirmation in Programmed Instruction." Journal of Educational Psychology, 1962, 53(6), pp. 250-253.

LaDuke, C. V. "The Measurement of Teaching Ability: Study No. 3." Journal of Experimental Education, 1945, 14, pp. 75-100.

Leitenberg, H.; Agras, W.; Thompson, L.; and Wright, D. "Feedback in Behavior Modification: An Experimental Analysis in Two Phobic Cases." Journal of Applied Behavior Analysis, 1968, 1, pp. 131-137.

Lessinger, L. M. "Accountability for Results: A Basic Challenge for America's Schools." Accountability in Education, Charles A. Jones, Publishing Co., Worthington, Ohio.

Lessinger, L. M., and Tyler, R. W. Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1971.

Lins, L. J. "The Prediction of Teaching Efficiency." Journal of Experimental Education, 1946, 15, pp. 2-60.

Levitt, T.; Schaff, M.; and Sayre, E. "The Use of Direct and Continuous Measurement to Evaluate Reading Materials and Pupil

Assessment." Strategies for Teaching Exceptional Children, Love Publishing Co., Denver, Colo., 1972, pp. 416-432.

Lublin, S. C. "Reinforcement, Schedules, Scholastic Aptitudes, Autonomy Need and Achievement in a Programmed Course." Journal of Educational Psychology, 1965, 56(6), pp. 195-302.

Lumsdaine, A. Some Conclusions Concerning Student Response and a Science of Instruction, Washington, D. C.: National Academy of Science and National Research Council, Publication No. 943, 1961.

MacPherson, S. J.; Dees, V.; and Grindley, C. G. "The Effect of Knowledge of Results on Learning and Performance II: Some Characteristics of Very Simple Skills, Quarterly Journal of Experimental Psychology, 1948-49, 1968-78.

Madsen, C.; Becker, W.; and Thomas, D. "Rules, Praise, and Ignoring Elements of Elementary Classroom Control." Journal of Applied Behavior Analysis, 1968, 1, pp. 139-150.

McKenzie, Hugh S.; Clark, M.; Wolf, M. M.; Kothera, R.; and Benson, C. "Behavior Modification of Children with Learning Disabilities as Tokens and Allowances as Back-Up Reinforcers." Exceptional Children, 1963, 34, pp. 745-753.

McNemar, Quinn. Psychological Statistics. John Wiley and Sons, Inc., New York, N.Y., 1969.

Medley, D. M., and Mitzel, H. "Some Behavioral Correlates of Teacher Effectiveness." The Journal of Educational Psychology, 1954, pp. 227-239.

Melaragno, R. J. "Effect of Negative Reinforcement in an Automated Teaching Setting." Psychological Reports, 1960, (7) pp. 381-384.

Melching, W. H. "PI Under a Feedback Schedule." NSPI Journal, 1966, 5(2), pp. 14-15.

Milliken, Wm. G. "Making the School System Accountable." Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1971, pp. 24-27.

Mitzel, H. E., and Medley, D. M. "Pupil Growth in Reading--An Index of Effectiveness." The Journal of Educational Psychology, 1954, pp. 227-239.

Moore, J., and Smith, W. L. "Role of Knowledge of Results in Programmed Instruction." Psychological Reports, 1964, 14, pp. 407-423.

Morsh, J. E.; Burgess, C. G.; and Smith, P. N. "Student Achievement as a Measure of Instructor Effectiveness." The Journal of Educational Psychology, 1953, pp. 79-88.

Moss, D. E., and Neidt, C. O. "Applicability of Information Theory to Learning." Psychological Reports, 1969, 24, pp. 471-478.

Novak, John, and Moser, Gene. "A Study of the Effect of Timed Pupil Feedback on the Teaching Behaviors of Biological Science Teachers." (Paper Presented at the National Association for Research in Science Teaching Meeting.) Chicago, Ill., April, 1972.

Nyquist, E. G. "Measuring Purposes and Effectiveness." Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1972.

Orleans, J. S.; Clarke, D.; Ostreicher, L.; and Standlee, L. "Some Preliminary Thoughts on the Criteria of Teacher Effectiveness." Journal of Educational Research, May 1952, Vol. XLV, No. 9, pp. 641-648.

Patterson, G. R., and Brodsky, G. "A Behavior Modification Program for a Child with Multiple Problem Behaviors." Journal of Child Psychology and Psychiatry, 1966, 7, pp. 277-295.

Peter, Lawrence J. Prescriptive Teaching. New York: McGraw-Hill, 1965.

Peterson, Russell W. "Accountability in Elementary and Secondary Education." Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1971, pp. 24-27.

Porter, J. W. "Objectives of Accountability." Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1971.

Ramp, Eugene A., and Hopkins, B. L. A New Direction for Education: Behavior Analysis, 1971, Vol. 1, University of Kansas, 1971.

Reddick, T. "Self-Accountability Can Make the Difference." Teacher Educator, v7, n7, Winter 71-72, pp. 111-113.

Ripple, R. E. "Comparison of the Effectiveness of a Programmed Test with Three Other Modes of Presentation." Psychological Reports, 1963, 12, pp. 227-237.

Risley, Todd R. "Behavior Modification: An Experimental-Therapeutic Endeavor." Paper prepared for Banff International Conference on Behavior Modification, April 1969.

Rosenstock, E. J.; Moore, J. W.; and Smith, W. I. "Effects of Several Schedules of Knowledge of Results on Mathematics Achievements, Psychological Reports, 1965, 17, pp. 535-541.

Rule, Sarah. "A Comparison of Three Different Types of Feedback on Teacher's Performance." Behavior Analysis and Education, 1972, The University of Kansas, 1972.

Science Research Association Arithmetic Series: Diagnostic Tests, Montgomery County Board of Education, Montgomery County, California, 1965.

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

Stephens, Thomas M. Directive Teaching of Children with Learning and Behavioral Handicaps, Charles E. Merrill Publishing Co., Columbus, Ohio, 1970.

Stephens, Thomas M. Implementing Behavioral Approaches in Elementary and Secondary Schools, Charles E. Merrill Publishing Co., Columbus, Ohio, 1974.

Swets; Millman; Fletcher; and Green. "Learning to Identify Non-verbal Sounds: An Application of a Computer as a Teaching Machine." Journal of the Accoustical Society of America, 1962, 34, (7), pp. 928-935.

Taber, J. L.; Glaser, R.; and Schasfer, H. H. Learning and Programmed Instruction, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1965.

Thomas, Don E. "Preliminary Findings on Self-Monitoring for Modifying Teaching Behaviors." A New Direction for Education: Behavior Analysis: 1971, The University of Kansas, 1971.

Thomas, D. R.; Becker, W. C.; and Armstrong, M. "Production and Elimination of Disruptive Classroom Behavior by Systematically Varying Teacher's Behavior." Journal of Applied Behavior Analysis, 1968, 1, pp. 35-45.

Tyler, R. W. "Accountability in Perspective." Accountability in Education, Charles A. Jones Publishing Co., Worthington, Ohio, 1971, pp, 1-6.

Vastbinder, S. D. The Effect of a Timer on Study Behavior of a Seventh-Grade Boy, unpublished study, The Ohio State University, 1972.

Vergason, Glenn. "Accountability in Special Education." Exceptional Children, February, 1973, 9, pp. 367-373.

Weiner, N. The Human Use of Human Beings, Boston: Houghton-Mifflin, 1950.

Williams, C. D. "The Elimination of Tantrum Behavior by Extinction Procedures." Journal of Abnormal and Social Psychology, 1959, 59, p. 269.

Wittrock, M. C., and Twelker, P. A. "Prompting and Feedback in the Learning, Retention and Transfer of Concepts." British Journal of Educational Psychology, 1964, 34, (1), pp. 10-18.

Wolf, M. M., and Risley, Todd R. "Reinforcement: Applied Research." Paper presented at the Annual Conference of the Learning Research and Development Center, University of Pittsburgh, June 1969.