R E P O R T R E S U M E S

ED 015 657 24 EM 004 053

MATHEMATICS TEACHING BEHAVIOR CHANGES MADE BY INTERMEDIATE GRADE TEACHERS DURING A 15-WEEK PERIOD OF INSTRUCTION BY EDUCATIONAL TELEVISION. FINAL REPORT.

BY- CORLE, CLYDE G.
PENNSYLVANIA STATE UNIV., UNIVERSITY PARK

REPORT NUMBER BR-6-1234 PUB DATE MAR 67
GRANT OEG-6-10-358
EDRS PRICE MF-$0.50 HC-$3.44 84P.

DESCRIPTORS- *OBSERVATION, *RATING SCALES, *EDUCATIONAL TELEVISION, *ELEMENTARY SCHOOL MATHEMATICS, *INSERVICE TEACHER EDUCATION, OSCAR (EM)

PURPOSE OF THIS STUDY WAS TO QUANTIFY BEHAVIOR CHANGES OF FOURTH, FIFTH, AND SIXTH GRADE MATH TEACHERS WHILE THEY WERE TAKING A 15-WEEK IN-SERVICE EDUCATIONAL TV COURSE, TAUGHT BY THE AUTHOR. 16 PAIRS OF TEACHERS WERE SELECTED FROM 6 DISTRICTS TO PARTICIPATE. 1 MEMBER OF EACH PAIR WAS CHOSEN BY LOT TO TAKE THE COURSE, AND THE OTHER AGREED NOT TO VIEW IT. 7 TRAINED OBSERVERS VISITED, ON A ROTATING BASIS, EACH OF THE 32 TEACHERS 30 TIMES, 7 BEFORE, AND 23 DURING THE COURSE. THEIR BEHAVIOR WAS RECORDED ON THE 8 SCALES OF THE OBSERVATION SCHEDULE AND RECORD (ELEMENTARY MATHEMATICS) AND ANALYZED BY ANALYSIS OF COVARIANCE. 2 OF THE SCALES SHOWED DIFFERENCES BETWEEN TEACHING STRATEGIES OF THE EXPERIMENTAL AND CONTROL TEACHERS. GAIN SCORES ON 2 ACHIEVEMENT TESTS SHOWED THE EXPERIMENTAL TEACHERS TO BE SIGNIFICANTLY HIGHER THAN THE CONTROL GROUP. THE AUTHOR CONCLUDED THAT THE 15-WEEK PERIOD WAS PROBABLY TOO SHORT TO DOCUMENT BEHAVIOR CHANGES, AND THAT THE 20-MINUTE OBSERVATION PERIOD MIGHT NOT HAVE PROVIDED ENOUGH TEACHER-PUPIL INTERACTION TO JUSTIFY USE OF THE OBSERVATION INSTRUMENT. (LH)

ED015657

A BR-6-1234
PA 24

MATHEMATICS TEACHING BEHAVIOR CHANGES MADE BY INTERMEDIATE GRADE TEACHERS DURING A 15-WEEK PERIOD OF INSTRUCTION BY EDUCATIONAL TELEVISION

~~Project No. P. 3619~~
Grant No. OE 6-10-358

EM004053

Clyde G. Corle

March, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Pennsylvania State University
University Park, Pennsylvania

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT
Project No. P. 3619
Grant No. OE 6-10-358

MATHEMATICS TEACHING BEHAVIOR CHANGES MADE BY INTERMEDIATE GRADE TEACHERS DURING A 15-WEEK PERIOD OF INSTRUCTION BY EDUCATIONAL TELEVISION

March, 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EM 004 053

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the six school districts who assisted in this study and to the 32 teachers who opened their classrooms to observation visits. He is also grateful to the seven graduate students who made the visits and reported the teaching behaviors of the teachers during almost a thousand classroom observations. The study by Dr. Crayton L. Buck of Oswego, New York, contributed immeasurably to this investigation. Dr. Buck prepared the OScAR (EM) and tested it for his doctoral research.

Station WPSX-TV and its staff cooperated in every way to produce a quality television program for teachers. The author is especially grateful to Marlowe Froke, Director of Division of Broadcasting, Arthur Albrecht, producer, and Charles R. Furman, director of the programs. Mrs. Shirley Terwilliger, the typist who prepared this manuscript deserves a special acknowledgement.

C.G.C.

LIST OF TABLES

I. INTRODUCTION

Background: Important changes are taking place in the teaching techniques and the content of elementary school mathematics. Deans (1963) believed these changes to be the result of (1) rapid advances in the knowledge of mathematics, (2) a need for better articulation between grade levels, (3) a need for better understanding of the structure of mathematics, and (4) a need for differentiation in mathematics programs to provide for children of varying ability.

This reconsideration of the elementary school mathematics program has brought at least two questions to the attention of elementary school principals and supervisors. (1) Are the teachers in their school district competent, both in content and methodology, to present a comprehensive program of mathematics? (2) Are the parents of the children in their schools informed about the changes which will be necessary before a new program can be established? Many administrators have taken steps to secure positive answers to these questions. They have provided in-service education for teachers; they have encouraged continuing education and summer school participation; and they have established evening classes in mathematics for parents.

Should a state university share the responsibility for upgrading instruction in elementary school mathematics? Just how does a university see the problem of educating thousands of parents? Obviously, any university would find it next to impossible to find enough help for the local schools if they all asked

for it at once. The supply of professional persons who can work with teachers is limited, and few universities could provide enough assistance if every district wished to hold its own workshops. The Continuing Education Staff of the Pennsylvania State University believed that a course designed for teachers and broadcast over educational television would reach teachers in every district within its viewing area. School districts who want an in-service program could then recommend such a course for the teachers in their school. Parents seeking information about changes in mathematics instruction could also audit the course and learn with the teachers.

The Pennsylvania State University operates Station WPSX-TV, an educational television station which serves 22 central Pennsylvania counties. Through the correspondence department of continuing education, the Department of Elementary Education of the College of Education offered over WPSX-TV El. Ed. 426, Teaching Modern Mathematics, a three-credit course emphasizing new methods and new content for elementary school mathematics. The course could be taken either for undergraduate or for graduate credit. This 15-week course was presented in 45 half-hour broadcasts, two each week for lectures a demonstrations and the third one for questions and answers. The correspondence department provided textbooks and a study guide. The students completed fifteen assignments, each consisting of two lessons per week from the study guide, and returned the assignments to the correspondence department for correction. The midterm and final examinations were

proctored by local school supervisors who returned the examination papers to correspondence officials for correction.

Related Studies: Television as a Medium for In-service Education. The value of television as a method of bringing in-service education to teachers has been supported in a number of investigations. Some of these researchers used commercial outlets, some reported educational channel broadcasts and some had operated on closed circuit transmission. Although several of the studies supported television as an effective method of teacher in-service education, this support was generally based upon two methods of evaluation: (1) achievement tests to measure the learner's growth in subject matter and (2) attitude scales to determine how the subjects felt about their experiences.

Hunt (1961) presented 15 half-hour television broadcasts on individualized reading methods over a commercial station in central Pennsylvania. He selected 213 elementary teachers, half of whom viewed the broadcasts and half who did not. He used three techniques to evaluate the effectiveness of the broadcasts on the teachers' classroom behavior: (1) four post-treatment observations of the teachers in the classroom, (2) a Teacher's Attitude toward Teaching Reading Scale, and (3) a Teacher's Self Report on Teaching Reading. He also measured parents' and pupils' attitudes towards reading.

Hunt's conclusions were as follows:

1. The experimental treatment positively affected the observed classroom performance of the viewing teachers.

2. The experimental treatment positively affected the teachers' attitudes toward individualized reading.

3. The experimental treatment had a significant effect on what teachers reported they were doing in teaching reading.

4. The experimental treatment had a significant positive effect on the attitudes of viewing parents.

5. The experimental treatment had no measurable non-chance effect on children's achievement in reading or on their attitudes toward reading.

Although Hunt's study supports the effectiveness of television broadcasts on in-service education, his dependence upon self-reporting instruments alone leaves the real outcomes of the study open to some question. The classroom observations were made after the broadcasts, and since there were no pre-treatment visits it was not possible to reach any objective conclusion that the teacher's behavior changed after viewing the programs.

DeVault, Houston, and Boyd (1962) experimented with approximately 100 elementary teachers in the Dallas, Texas, area. Using content prepared by a research team, the researchers presented their in-service education ideas to teachers in four ways: (1) They broadcast 24 half-hour television programs of in-service mathematics instruction over an educational television station, (2) The same person who presented the television broadcasts gave face-to-face lecture-discussion - 12 hour-and-a-half sessions every other week, (3) They offered television with consultant services, in which the viewing teacher received visits (5.44 per teacher) and assistance with her teaching, and (4) They provided face-to-face lecture-discussion with consultant (4.95 visits per teacher) service.

The researchers used three achievement tests to measure improvement in teacher understanding of mathematics and teacher understanding of methods. The conclusions of this study were as follows:

> First, television was as effective as face-to-face lecture-discussion in changing the mathematics and methods understandings of teachers, in the reactions of teachers to the in-service education program, in changing all but one of nine components of the classroom practices of teachers, and in changing the mathematics achievement and interest of pupils in classes of participating teachers. Second, consultant services as a supplement to television and face-to-face lecture-discussion made a significant contribution in some situations. That is, consultant services resulted in more favorable reactions from television teachers; they resulted in greater mathematics achievement among heterogeneous pupils; and they resulted in greater general mathematics achievement of pupils of television teachers but not of pupils of face-to-face lecture-discussion teachers.
>
> Thus, hypothesis number one: There is no difference in the effectiveness of television and face-to-face lecture-discussion as media for in-service education of intermediate grade teachers of mathematics - is accepted.
>
> Hypothesis number two: There is no difference in the effectiveness of in-service education program presented by television or face-to-face lecture-discussion which are supplemented with consultant services as compared to those in-service education program without such consultant services - is rejected.

The first conclusion by DeVault and others (1962) cannot be accepted as stated above. Hypothesis number one states only that there was no difference in the effectiveness of television and face-to-face lectures-discussion as media for in-service education of intermediate grade teachers of mathematics. To say that one is as effective as another overlooks a number of other variables, any of which might be operating to influence the behavior of teachers.

The researchers probably should have reported only that the two sets of scores were not different from each other.

The results of this study, however, were measured by achievement tests of teacher and pupil knowledge. Any changes in teaching behavior which might have resulted from the various treatments used with these teachers were not reported.

Wittich (1961) studied the effectiveness of a series of 42 kinescoped television programs each 28 1/2 minutes long, designed for in-service teacher education to promote attitudes favorable to audiovisuals and to improve teaching practices in the use of audio-visuals. Wittich used four specially-constructed scales to determine the effect on actual day-to-day classroom practices, shifts in teachers' attitudes toward AV materials and techniques, teachers' opinion of the course, and the impact of the course on the teacher for whom it was intended.

Wittich reached the following conclusions:

1. The AV - TV course exerted a small but significant positive effect on AV utilization in the classroom.

2. Although the teachers initially held a favorable attitude toward all the AV concepts listed, a post-course scale showed improvement in attitude toward several of these concepts.

3. The teachers' comment on the AV - TV course was favorable to the course.

Although one can agree that the conclusions of this investigation are justified by the nature of the treatment, the study failed to produce evidence that teachers made any changes in their own classroom use of audio-visual materials.

Garry and Mauriello (1960) divided 45 fifth grade teachers whose classes were about to receive television instruction in French into two categories: (1) fluent and (2) non-fluent speakers of French. These 45 teachers were randomly assigned to four different treatment groups:

1. Those who viewed a televised teacher training program on French instruction and used tape recordings as practice materials with their classes.

2. Those who viewed the television programs, but who themselves planned the practice sessions with their classes.

3. Those who practiced with tape recordings to improve their own fluency and used tape recordings with their classes.

4. Those who practiced with tape recordings to improve their own fluency and planned their own practice sessions with their classes.

Specially constructed listening tests were used to measure the results. The researchers reached the conclusion that the method of presentation to the teachers made little difference in the level of skill and understanding they reached. The researchers reported after a year's study that television instruction alone without appropriate follow-up practice is largely ineffective.

To determine the relative effectiveness of four methods of in-service education in the use of radio and television in the classroom, Glasgow (no date given on publication) worked with 181 elementary and secondary teachers and their pupils to determine which of the following methods gave greatest support to the broadcasts:

1. Workshops for teachers
2. Supervisory visits
3. Use of special printed material
4. Use of films and recorded tapes

Glasgow found that none of the four investigated factors proved superior to any of the others or to any combination of them. The teachers expressed a preference for the workshop method but they also recognized a need for a diversified method of in-service instruction.

The researcher used pupil-achievement tests and teacher-opinion surveys as evidence of program effectiveness. There was no systematic observation of changes in teacher behavior resulting from the different treatments.

Follis (no date given on publication) sought answers to four questions about the use of closed-circuit television for making observations of pupils in the classroom.

1. As viewed over closed-circuit television, did the classroom seem to be a natural setting?

2. Were the audio and video systems effective in providing the emotional feel of the classroom?

3. Were the viewing sessions conducive to clear observation and good discussion?

4. Were the television viewing sessions a valuable additional medium for in-service education?

The subjects for the study were 20 principals and educators, 20 teachers, and two professors of education. The consensus of these persons was favorable to the potential of television as a method of observing the classroom behavior of teachers and students. However, the researcher reported only the teachers' opinions as the bases for his evaluations.

Frazier and Evans (1960) presented ten half-hour telecasts on elementary science topics approved by 151 teachers of third and

fourth grade classes. The purpose of the study was to explore the effectiveness of the television programs in increasing teacher understanding of elementary science and in stimulating pupil activity in this field.

The investigators reached the following conclusions:

1. Teachers significantly increased their feelings of competence in teaching science in the classroom.

2. Teachers expressed a willingness to accept such help in other curricular areas.

3. Teachers evaluated the programs as having a high pupil interest and reported significant growth in problem solving behavior.

4. Pupil growth as measured by use of a standardized test was not significant.

The studies cited in the foregoing discussion appear to support a wider use of television for the in-service education of teachers. A number of them report significantly better performances on the part of teachers who have taken part. With the exception of Hunt (1961), however, the investigators used only test results, rating scales, questionnaires or opinion samples to reach their conclusions about television as an in-service medium. The true measure of a teacher's effectiveness is how that teacher performs in the classroom, and the only way to study a teacher's performance is to make systematic observations of day-to-day lesson presentations.

Systematic observations of teacher behaviors require a special kind of instrument for recording and quantifying teacher-pupil interactions. Such techniques had been in use for a number of years, and a number of published studies could be found which offered

assistance to anyone wishing to study teacher behaviors. (See pp. 12f) For example, from 1955 to the early 1960's, Medley and Mitzel (1963) had developed and standardized classroom observation instruments. Such instruments might have been employed if any of the researchers had been interested in behavior changes in the classrooms of teachers involved in their studies.

Although Hunt (1961) did make classroom observations, all of his visits took place after the teachers had received the television treatment. Hunt did not make use of a systematic observation schedule; he reported no visits prior to the treatment period; and he offered no objective evidence that changes actually occurred.

Purpose of the Study: The foregoing review of studies related to educating teachers by television broadcasts has left unanswered a persistent and very critical question. Will teachers change their mathematics teaching behavior while they are taking a college-credit television course in the teaching of elementary mathematics? If this question could be answered affirmatively, educational television services then should be extended to provide such courses for all teachers in a viewing area. If, however, no changes can be identified, then those who direct program planning should exercise caution in scheduling broadcasts of this nature.

Statement of the Problem: The investigator undertook this study to secure new information about teachers and the circumstances under which they change their teaching behaviors. The specific problem to be researched was:

Will changes in the mathematics teaching behavior of intermediate grade teachers recorded while they are taking a fifteen-week television course in teaching mathematics differ from the changes made by other intermediate teachers who were not taking such a course?

II. METHOD

Design for the Study. The investigator assisted Buck (1967) in selecting a number of teacher activities which might occur while a teacher was instructing an intermediate grade mathematics class. He included motivating strategies, contributions by teachers, and learner reactions, both verbal and physical. From these elements of classroom behavior Buck prepared an evaluative instrument which he named the OScAR (EM). (Observation Schedule and Record, Elementary Mathematics) The purpose of the OScAR (EM) was to permit an observer to visit a mathematics class and to leave it with an objective report of the teaching behaviors which occurred during the visits.

Any study of behavior changes requires an evaluation of pre-treatment, or threshold behaviors as well as an appraisal of those which might result from the treatment. Trained observers, using the OScAR (EM) made seven threshold visits to the experimental teachers (those viewing television lessons) and 23 visits to the same teachers during the viewing period (after the broadcasts began). An equal number of threshold visits were made to control teachers (those not viewing televisions lessons) and each control teacher was visited 23 times during the broadcasts.

The television course was Elementary Education 426. Teaching Modern Mathematics, taught by the investigator on Station WPSX - TV

from February 8 to May 27, 1966. The course consisted of instruction in methods of teaching mathematics to elementary school pupils.

The observers who visited the classrooms included seven experienced persons: teachers, supervisors, administrators and college instructors who had received training in the use of the OScAR (EM). Their visits consisted of a scheduled twenty-minute period during a regular mathematics class.

Preliminary work with the OScAR (EM) indicated that teaching behaviors in intermediate grade methematics classes can be classified into eight different kinds of teacher-pupil interaction. The OScAR was therefore divided into eight scales, each of which appeared to measure independent elements of teacher performance.

Threshold observation visits to teachers who were classified experimental and control began several weeks before the television course broadcasts began. The observers continued to make observations of both groups of teachers after the experimental teachers had begun working on the lessons. The control teachers did not view the broadcasts or in any way take part in the television course.

In brief, the design of the study was as follows:

1. To follow up a course in elementary school mathematics methods which was presented to teachers by educational television.

2. To prepare an observation schedule and record for reporting the mathematics teaching behavior of intermediate grade teachers.

3. To measure, before the lessons began, the teaching behavior of two sets of teachers, one set who would take the course and the other who would not.

4. To measure, during the viewing period, the classroom teaching behaviors of both sets of teachers.

5. To determine whether changes in the teaching behaviors of the viewing group were different from changes in the non-viewing group of teachers.

The Development of an Observation Instrument. It is very difficult to obtain quantitative data about teacher behavior without using some device such as a rating form or a check list. Although such forms have been used in other teacher evaluation studies, the investigator could not find an instrument which could be used to collect data about intermediate grade teachers when observers visited their mathematics classes. Early studies in the development of an observation schedule and record proved helpful, but these ideas were not usable until they were adapted to mathematics instruction in the intermediate grades. Much of this adaptation was accomplished by Buck (1967) as he worked on a parallel doctoral study.

Studies Related to Observation of Teachers: The early history of an observation schedule for studying teacher behavior goes back to the early 1900's. Horn (1914), Barr (1929) and Wrightstone (1934) reported some early studies in making observations of a teacher's behavior in the classroom. These studies attempted to identify specific activities of teachers and to provide a check list for an observer to use in reporting their occurrence.

Thomas (1929), Withall (1949), Cornell and others (1952) and Hughes (1959) attempted to classify such observations and establish identifying terminology for the interaction of teachers

and pupils in a learning situation. Withall called this interaction the classroom climate, and to identify the behaviors more explicitly he developed a set of seven categories for teacher utterances. He then could classify observed behaviors according to the set of categories he had devised.

Medley and Mitzel (1963) constructed an instrument which they called the OScAR (Observation Schedule and Record). They used this instrument with teacher education graduates to provide quantitative data regarding the behavior of beginning teachers. The OScAR enumerates representative items of behavior within certain categories. An observer simply checks any item on the list if that behavior occurs during an observation period. The observations were timed with five-minute intervals being spent in checking the items in each of several categories. This system makes it possible to take two important steps in the process of measuring classroom behavior: (1) to secure a record of the sample of behaviors to be measured and (2) to quantify that record for further statistical study.

The original OScAR represented an adaptation of ideas about general observations of classroom behaviors reported in earlier studies. The designers of the instrument, however, had not overlooked the possibility of preparing such a schedule for a particular subject area. Schueler, Gold, and Mitzel (1962) raised the question of whether further development should be in the direction of a common schedule for all teaching or whether particular subjects need detailed, individual sections.

Other researchers have worked with the OScAR, each to achieve a somewhat different purpose. Bowers and Soar (1961) used it to compare teachers in the area of human relationships. Wilk and Edson (1962) attempted to use it in a predictive study of teacher behavior. Schueler, Gold, and Mitzel (1962) used the OScAR to evaluate kinescopes of student teacher behavior.

Wright (1959) believed such an instrument could be used to measure the verbal teaching behaviors of teachers in secondary mathematics. Wright's model classifies content process and attitude by the level of the rigor and the pupil participation in the lesson. She found the instrument useful in discriminating between various groups and between styles of teaching.

Buck, (1967) wished to use an observation schedule and record to study differences in intermediate grade teaching behaviors when years of experience and mathematics achievement test scores were used to classify teachers. Buck needed an instrument especially designed to study the mathematics teaching activities of teachers, and he prepared a modification of the OScAR which he called the OScAR (EM). He then selected fourteen pairs of intermediate grade teachers, seven pairs with fifteen or more years of experience and seven pairs who had taught five years or less. Each pair of teachers consisted of a teacher whose mathematics test scores were high and one whose scores were low. The pairs of teachers in Buck's study taught either in the same building or under the supervision of the same person. Buck made ten visits to each teacher, using

the OScAR (EM) to record his impressions of their mathematics teaching behavior.

The observers' scores from the 280 visits supplied the data with which Buck tested and evaluated the elementary mathematics version of the OScAR. After his visits to these fourteen pairs of teachers Buck concluded:

1. When their scores on mathematics achievement tests were high, teachers scored higher on the eight scales of the OScAR (EM). Although Buck associated no qualitative evaluation of teacher behavior with higher scores on the tests, the items on most of the scales represented recommended teaching practices.

2. When years of experience was a factor, the teachers' scores on the OScAR (EM) were not significantly different.

3. When years of experience and scores on achievement tests were combined, there was no significant difference in the OScAR (EM) scores.

Since Buck's study was so closely related to this investigation, a detailed account of his work in developing the OScAR (EM) will be included in this report (Appendix B).

Buck's preliminary study of the OScAR (EM) made several important contributions to this investigation: (1) It tested the instrument on a group of teachers who were judged to be a sample from the same population as those teachers taking part in this study of television instruction, (2) It established scales which could be used to identify specific kinds of classroom teaching behavior, (3) It provided evidence of the reliability of each scale; (4) It demonstrated that the scales were operating independently, and (5) It provided evidence that the scales discriminated between teachers.

The limitations of any study which uses an observation schedule and record as a method of gathering data about teacher behavior have been taken into consideration. Boyd and DeVault (1966) reviewed the literature on the observation and recording of behavior which has been published during the last six years.

The reviewers pointed out some of the weaknesses of the descriptive category type of instrument.

> "In strict terms descriptive categories are based solely on non-interpretative descriptions of observable overt behavior; the interpretative and inferential stages are made subsequent to the collection of the data. Evaluation datum categories require various degrees of interpretation at the time the data are collected; inferential analysis follows. It may be argued that interpretations and inferences made on overt data by someone not present at the time the behavior occurred may distort and/or omit aspects of the situation....To require the observer to record all behaviors is a monumental task. Yet to have the observer do less, when employing purely descriptive categories, may result in serious misrepresentation of the observed situation....The use of only overt behavior would remove, in addition, the concept of purposiveness of human behavior....It is not always possible to classify a unit of behavior into a particular category without considering the sequence of immediate acts or units...
>
> "A variable defined in terms of descriptive behaviors is based on the proposition that the descriptive behaviors are as complete as possible. This is an essential point that needs to be examined carefully....If a behavior is to be observed it should not be so general that its meaning is lost because the observer could not record the conditions under which it occurred.
>
> "If the researcher wants to treat his observations in quantitative terms, he is faced with the problem of measurement....It is not uncommon to find units defined in terms of both time and acts....The advantage of time units is that a sequential analysis of total group interaction can be readily made....If the unit is the act, then it is first necessary to define the act.

"The emphasis on verbal behavior...reflects the heavy emphasis placed on verbal behavior in classroom learning situations....Withall (1949, p. 349) stated, 'A teacher's verbal behavior is assumed to represent adequately her total behavior.' It would seem reasonable to assume that the burden of proof (of the above statement) rests with the researcher. Serious doubts about the above assumption would be raised by psychiatrists and social psychologists.

"The proper focus of the observer, that is, the position from which the observer should perceive the act has been an intriguing and vexing issue in observational data collecting....It is possible for the person being observed to have an influence on the observer...; the fixed roles given the observer may result in the recording of very bizarre observations if the observer is unable to interpret these types of behavior in terms of the interpersonal dynamics of the situation."

Boyd and DeVault have raised many questions which concerned the investigator as the study developed. The need for teacher-behavior studies in the elementary schools, however, is a pressing one, and regardless of the limitations, which are indisputable, researchers must continue to try to find workable methods of observing and reporting classroom strategies.

The Selection of the Teachers for the Study. WPSX-TV serves 22 counties in central Pennsylvania, and educational programs broadcast to the schools are supervised by the Allegheny Educational Broadcast Council (AEBC). From the school districts listed as participating members of AEBC, six districts took part in the study. The districts were arranged in order by using a table of random numbers, and with two exceptions the first six districts on the ordered list were chosen.

1. School districts in which the investigator for this study had previously conducted a series of mathematics workshops were eliminated.

2. Any district not wishing to take part in the study was excused and the next district on the selection list was invited.

Three pairs of intermediate grade teachers were selected from each school district. In almost every case both teachers in each pair taught in the same building. When such an arrangement was not possible both teachers in the pair were supervised by the same person. All teachers taught fourth, fifth or sixth grade classes. One teacher from each pair was chosen by lot to take the El. Ed. 426 course in teaching mathematics. The other member of the pair agreed not to view any of the lesson presentations or study any of the course material during the experimental period. Those teachers who took the television course were designated experimental teachers and those who did not take the course control teachers.

Measuring Teacher Achievement. Before the television broadcasts began, all experimental and control teachers took two tests of mathematics achievement. The first test was the Test of Arithmetic Principles by Stoneking and Welch (1961). The authors reported a reliability coefficient of .92, but they gave no explanation of the formula they used to compute that statistic. (Page 25) The second test was A Test of Mathematics Vocabulary by Corle (1966). The reliability coefficient for this test (.89) was computed by the Kuder-Richardson formula #20 when the test was used with in-service teachers.

Following the broadcasts, the experimental and the control teachers again took the two achievement tests.

Classroom Visits. Seven trained observers made 30 regular scheduled visits to each of the 32 teachers involved in the study. The sixteen teachers in the experimental, or television viewing group, were visited seven times before the television course opened and 23 times after the broadcasts began. Visits to the control, or non-viewing teachers, corresponded exactly to those made to the experimental group. The observers followed essentially Buck's (1967) procedures.

1. Times for visits were arranged to enable observers to complete their observation schedules.

2. Teachers agreed not to give tests or study periods during the observation time.

3. The actual observation time for each visit was 20 minutes, five minutes each for the four sections of the test.

4. The observers coded the OScAR (EM), so teachers could not know which behaviors were being recorded.

5. The seven observers rotated among the 32 teachers to minimize a "halo effect" from repeated observations.

6. There was no feedback to teachers to indicate what kinds of behavior the observers were recording.

7. The OScAR (EM) forms, when completed, were scored by the use of Buck's eight scales.

The Television Course. El. Ed. 426 Teaching Modern Mathematics was a graduate level course for elementary teachers (See page 2). It was patterned after a course carrying the same name and number which is offered both on the University campus and in University continuing education centers throughout the commonwealth. The first two lessons consisted of a philosophical and methodological treatment of the changes in emphasis in elementary school mathe-

matics. The remaining broadcasts each featured some mathematical topic, and the instructor stressed new ways to present each topic. Concrete objects, charts, and a wide variety of teaching aids featured each broadcast. A special effort was made to keep these materials simple and easy to make. The instructor made no effort to indoctrinate teachers or to suggest that they follow his example. He assumed that if such devices enabled teachers to learn more readily, the teachers might use them in their own classrooms.

The instructor used his own textbook (Corle, 1964) and the lesson units followed the textbook material. The textbook recommends considerable work with manipulative materials and offers many illustrations of simple devices for pupil and teacher use. Both the television broadcasts and the textbook stressed teacher-pupil interaction, problem solving through a discussion technique, and guided discovery as a desirable teaching procedure. A short summary of the course follows:

1. Teachers must learn to use a standard mathematical vocabulary which applies to a topic regardless of the grade level being taught.

2. Elementary school mathematics must be enlarged to include ideas from related areas of mathematical knowledge. Geometry, algebra, sets, logic, and similar topics are all within the understanding of elementary children and use of these topics expand and enrich their mathematical experiences.

3. Every computational operation which is taught in the elementary schools is understandable and every pupil must be given an opportunity to learn not only how the operation is performed but why it is done that way.

4. Guided discovery is an effective method of presenting mathematical ideas to children.

The observation schedule and record used by the observers in this study was oriented to the above four principles of teaching modern mathematics to elementary school children.

The Nature of the Data. The seven observers made 960 visits to the 32 teachers taking part in the study. The experimental teachers were visited 480 times and the control teachers 480 times. There were 112 threshold visits in each group and 368 visits made after the television lessons began. The observers computed the teacher behavior scores by counting the tallies for all of the items on each of Buck's eight OScAR (EM) scales. If a behavior occurred during a visit, a tally was entered for the appropriate item. The observer tallied the item only once during a visit, no matter how often he saw it. Thus each visit to a teacher produced eight scale scores, and the classroom teaching behavior of each of the 32 subjects could be identified by 30 x 8, or 240 scale scores.

Each of the 32 teachers took a pre- and post-treatment administration of two achievement tests (See page 19). Scores on each tests were computed by counting the number of correct items. These tests measured (1) understanding of basic mathematical principles used in the elementary school and (2) familiarity with the mathematical vocabulary appropriate to elementary school use.

The Treatment of the Data. This study attempted to provide an answer to the following problem. Will changes in the mathematics

teaching behavior of intermediate grade teachers, reported by observers while the teachers are taking a 15-week television course in teaching mathematics differ from that of other intermediate grade teachers not taking the course?

The investigator sought to answer this question by using the following procedures:

(I) He used analysis of covariance with teachers as a nested factor to study differences in means when the total scores on all of the scales were added together.

(2) He then used analysis of covariance with teachers as a nested factor to study differences in the means of each scale, computed separately.

III. RESULTS

Statistical Analysis of the Observation Data. Table 1 shows the data which was developed by submitting the total scores on the OScAR (EM) to analysis of covariance using teachers as a nested factor.

Table 1. COVARIANCE ANALYSIS
ALL OBSERVATION SCHEDULE SCORES

Source	SS_x	SS_y	S_p	ADj ssy	DF	Mean Square	F ratio
Among Teachers	89,166	32754.	27,276	26,145	31	843.4	<1
Between Groups	.0039	588.	1.52	587	1	587	<1
Between Teachers Within Groups	89,166	32,166	29,274	25.558	29	881.3	
Scales	343,014	378,593	347,017	160,826	7	22,975	40.66**
Groups X Scales	3,189	5,271	1,644	4,896	7	699	1.24
Among Teachers Within Scales	687,331	508,305	392,614		224		
Between Teachers Within Groups Within Scales	341,128	124,442	43,953	118,779	209	565	

**Significant >.01

Table 1 shows that there were no significant differences either among the teachers or between the groups. This fact indicates that the observed behaviors reported from the visits were much alike, no matter what teacher was involved. Groups had been defined according to the treatment they received, television or no television. There was no significant difference between the groups when total scores on the eight scales were analyzed.

Table 1, however, did show a significant difference within the scales. Since each of the scales was developed as an independent measure, a high F ratio (40.66) between the mean scores was expected. Although the difference between groups by scales was not significant when total scores were considered, there was a possibility that differences between groups actually had occurred if each of the scales were considered independently. The F ratio of 1.24, between the mean scores when groups were combined with scales, was not significant. If, however, offsetting differences in scores occurred when scales were used to measure teacher behaviors, there might be differences between the means of one or more of the scales. The mean scores on each of the eight scales were then examined through analysis of covariance using teachers as a nested factor. This was done to find out whether scale differences between groups were significant. Table 2 shows these data.

Table 2
Covariance Analysis of OScAR (EM) Scores by Scales

Scale	Experimental (TV) Pre-test Mean (x)	Post Test Mean (y)	Adjusted Mean	rxy	Control (No TV) Pre-test Mean (x)	Post Test Mean (y)	Adjusted Mean	rxy	F ratio
1	124.69	124.86	124.86	.26	126.75	106.63	106.45	.26	2.50*
2	42.00	28.75	27.91	.49	35.81	31.50	32.33	.49	< 1
3	116.00	116.37	115.49	.34	108.00	115.56	116.45	.34	< 1
4	96.38	121.19	121.38	.18	100.00	116.63	116.44	.18	< 1
5	48.25	53.31	53.46	.20	49.94	45.75	45.61	.20	< 1
6	66.94	53.31	52.56	-.05	82.13	69.00	69.75	-.05	5.21*
7	59.88	68.50	67.80	.35	54.50	61.00	61.70	.35	< 1
8	12.63	16.06	15.14	.59	8.94	11.88	12.80	.59	< 1

* F ratio 4.17 significant at .05 level of confidence.

Table 2 shows that the adjusted mean of experimental teachers was somewhat higher than that of control teachers. The difference, however, was not significant. Scale 1, Teacher Directing Strategies and Pupil Responses, consisted of the following items: (The items for all scales appear in Appendix B)

1. Pupil agrees with teacher.
2. Pupil offers hypothesis.
3. Teacher poses leading or structuring questions.
4. Teacher rephrases or restructures pupil contribution.
5. Teacher requests pupil to repeat contribution.
6. Teacher applies logic to mathematical ideas.
7. Teacher introduces concept through problem.
8. Teacher uses illustration.
9. Teacher uses social application.

The items on Scale 1 include teaching strategies which, when used in a mathematics class, would produce a desirable kind of lesson environment for elementary school children. Even though the difference was not significant, the observers apparently saw the teachers in the viewing group using Scale 1 strategies more often than the non-viewing teachers. If such were the case, the motivation to use these strategies might have come from the telecasts.

Scale 2, Pupil-Teacher Interaction, consisted of the following four items:

1. Pupil gives verbal assistance to another pupil.
2. Pupil expresses disagreement with teacher.

3. Teacher allows pupil choice of solution.

4. Teacher interrelates mathematical ideas.

There was no significant difference between the mean score of teachers in the experimental group and that of teachers in the control group when measured by Scale 2.

Scale 3, Teacher Vocabulary Emphasis, contained seven items, as follows:

1. Decimal
2. Number
3. Numeral
4. Zero
5. Divide
6. Regrouping
7. Teacher uses Blackboard

The mean score of the experimental teachers was not significantly different from that of control teachers on Scale 3.

The six items on Scale 4, Classroom Discussion Strategies, are shown below:

1. Pupil offers possible solution.
2. Pupil volunteers pertinent information.
3. Teacher asks pupil to illustrate answer.
4. Teacher locates key contributor and calls on him.
5. Teacher waits for pupil to formulate answer.
6. Teacher uses mathematical rule or principle.

There was no significant difference between mean scores of experimental and control teachers on Scale 4 of the OScAR (EM).

The six items on Scale 5, Teacher Motivating Behavior, are shown below:

1. Sum
2. Teacher uses environmental object.
3. Teacher encourages alternate solution.
4. Teacher adds enrichment for pupils.
5. Teacher develops rule or formula for procedure.
6. Teacher questions correct answer.

The mean score of teachers in the experimental group was not significantly different from that of the control group on Scale 5 of the OScAR (EM).

Scale 6 contained five items as follows:

1. Pupils compute on chalkboard.
2. Teacher asks pupil how he solved problem.
3. Teacher refers statement or question to another pupil.
4. Teacher makes reproving remark.
5. Teacher accepts solution without comment.

The mean score for the control teachers on Scale 6 of the OScAR (EM) was significantly higher than that of the experimental teachers. The teacher activities listed under Scale 6 could well be classified as undesirable methodology, much as Scale 1 seemed to reflect a more desirable approach. For example; (1) having the pupil put the example on the blackboard, (2) asking for an explanation of his

work, (3) referring his report to classmates for agreement, (4) teacher reproving a pupil when pupil is wrong and (5) accepting a correct answer without comment all seem to be typical of undesirable instructional practices.

This study was not intended to be a qualitative evaluation of teachers. However, the only significant difference between groups appeared on Scale 6. Since the mean score of the control group exceeded that of the experimental group, the observers saw control teachers using these items more often than they saw experimental teachers using them. The television course could not be responsible for increasing the frequency of any practice among control teachers, but these lessons might have reduced somewhat the use of such procedures by the experimental teachers. The television lessons advocated a pupil-centered, discovery type of lesson, utilizing skills like those indicated in Scale 1. After viewing the broadcasts and completing the written lessons, some of the teachers possibly substituted other practices for some of those listed in Scale 6.

Scale 7, Teacher Vocabulary Reference to Fundamental Operations, was composed of four items.

1. Add
2. Multiply
3. Remainder
4. Subtract

There was no difference in the mean score of experimental teachers on Scale 7 of the OScAR (EM) from that of the control teachers.

Scale 8, Information Seeking by Pupils, contained only two items:

1. Pupil requests more information.
2. Pupil seeks teacher's help.

There was no difference in the mean scores of experimental and control teachers on Scale 8 of the OScAR (EM).

The mean scores of experimental and control teachers on Scales 2, 3, 4, 5, 7 and 8 did not differ significantly from each other. Some difference, although it was not significant, occurred in Scale 1. Observers reported a higher frequency of desirable behaviors among the experimental, or television viewing teachers. On Scale 6, the control, or non-viewing teachers made significantly higher scores than the viewing teachers. The items in Scale 6 represented doubtful or undesirable practices. Perhaps the television viewing teachers found some ideas either in the telecasts, the textbook, or the study guide which led them to make these corrections in teaching strategy.

Teacher Achievement in Mathematics: The achievement tests given to the teachers in both treatment groups were discussed on Page 19. All teachers in both groups took the tests before the threshold visits began and again when the television viewing period ended. Table 16 (Appendix C) shows the teachers' scores on the tests.

Table 3 shows the data on the Test of Arithmetic Principles for the experimental and the control teachers on the pre- and post-test.

Table 3
Pre- and Post-test Results of Experimental and Control Teachers on the Test of Arithmetic Principles

Teachers	Pre Test Mean	Post Test Mean	Adjusted Mean	Mean Square	F-Ratio
Experimental (TV)	54.00	57.75	56.41	131.76	
Control (No TV)	50.00	50.94	52.28	30.11	4.37*

*Significant at the .05 level of confidence

The experimental teachers made significantly higher gains on the Test of Arithmetic Principles. Teachers who viewed the broadcasts and studied the written lessons improved their knowledge of arithmetic principles during the viewing period.

Table 4 shows the pre- and post-test data on the Test of Mathematics Vocabulary.

Table 4
Pre- and Post-test Results of Experimental and Control Teachers on the Test of Mathematics Vocabulary

Teachers	Pre Test Mean	Post Test Mean	Adjusted Mean	Mean Square	F-Ratio
Experimental (TV)	35.13	47.19	45.04	530.72	
Control (No TV)	29.44	34.31	36.46	33.58	15.80**

**Significant at the .01 level of confidence

The experimental teachers also made significantly higher gains on the Test of Mathematics Vocabulary. The teachers who viewed the broadcasts and completed the lesson assignments appeared to improve in their ability to understand mathematical vocabulary.

The foregoing data indicates that the teachers who studied the course by television achieved a higher level of mathematics understanding than those who did not. All of the experimental teachers submitted their lessons promptly, took the midterm and final examination at the proper time and often sought special help with their course work. The telecasts of regular lessons were repeated, and teachers often reported that they had watched both showings. The viewing teachers in a school district sometimes organized a kind of seminar to watch the broadcasts. After each showing they discussed the ideas presented and completed their assignment sheets together. A combination of telecasts, readings, lesson sheets, and discussion apparently raised the level of mathematics achievement for these teachers.

Summary of the Data: The purpose of this study was to determine whether changes made in intermediate grade teachers' mathematics teaching behaviors while they were taking a television course in mathematics teaching methods differed from changes made by other intermediate teachers not taking the course. The data appears to support the following statements:

1. There were no significant differences among the teachers when total scores on the eight scales were considered.

2. Using total scores on the eight scales, the difference between groups (television and no television) was not significant.

3. Differences between the scales were significant when total scores on all scales were analyzed.

4. Differences between scales within groups were not significant when the total scores on all scales were analyzed.

When each scale was analyzed separately, the following statements could be supported:

5. On Scales 2, 3, 4, 5, 7, and 8 there was no significant difference between groups.

6. On Scale 1, although the difference was not significant, there was some indication of more desirable teaching behavior among the television viewing group of teachers.

7. On Scale 6, there was a significant difference in mean scores of the two groups. The television viewing teachers appeared to make desirable changes in their teaching strategies more frequently than the control teachers.

Covariance analysis of achievement test scores showed the following changes.

8. The experimental teachers (television viewing) made significantly greater gains on the Test of Arithmetic Principles.

9. The experimental teachers made significantly greater gains on the Test of Mathematics Vocabulary.

IV. DISCUSSION

Distribution of Teaching Time During the Visits. An unforeseen element of teacher behavior became apparent as the observers completed the early rounds of visits, and that element was the low percentage of time spent on developmental lessons. The OScAR (EM) was designed to measure teacher behavior in a verbally interacting classroom atmosphere. When it became apparent to the observers that the teachers were spending much less than the recommended portion of the class period in developmental teaching, the observers decided to keep a record of the time spent in various kinds of class activity. Recent studies by Shipp and Deer, (1960) Zahn, (1966) and Shuster and Pigge, (1965) agreed that teachers should spend at least 50% of the class time in teacher-pupil verbal interaction, developing new ideas, and applying them. Buck (1967) reported that about 12 percent of the teachers' time was devoted to preparing the class for its lesson, 33 percent to developmental teaching, 38 percent to review and drill, and 17 percent to correcting assignments. Milgram's (1967) estimates were as follows: going over previous assignment, 25 percent; oral or written drill, 51 percent; introduction of new concepts or developmental activities, 23 percent; unrelated activities, 1 percent. Both observers reported that there appeared to be no difference between treatment groups in the percentage of time spent in the various kinds of teaching activity.

A characteristic lesson, as reported unofficially by these observers, consisted of an assignment from the book, a brief explanation by the teacher of how to solve the examples and then a supervised work period. The next class began by assigning various pupils an example to put on the blackboard, a recitation by the pupil of how he solved the example, and corrections made by the teacher. Another assignment, very much like the previous one, generally followed the correction of homework papers.

Since the purpose of the study was to determine any changes which might be attributable to television instruction, there could be no feedback from the observers to the teachers. Teachers were unaware of the nature of the behaviors being observed, and therefore found little motivation to change the daily routines.

Obviously, the greater emphasis on drill activities and correcting assignments reduced the number of item responses to the OScAR (EM) both for the experimental and the control teachers.

The experiment was conducted in a geographical area which has been slow to accept the teaching behaviors commonly identified with the so-called modern approach to elementary school mathematics. Unofficial reports from the observers indicated that many of the teachers felt insecure in their knowledge of the modern treatment of mathematics and somewhat uncomfortable in being visited by "experts". It is possible that these teachers utilized teaching procedures which would expose them as little as possible to any

kind of critical evaluation. Perhaps on days they were not visited they made greater use of pupil-centered methods.

Mathematics Achievement: The teachers who took the television course gained significantly more mathematical knowledge than those teachers who did not take any course during the experimental period. The test scores of the experimental teachers were higher both on the test of arithmetic principles and on the test of mathematics vocabulary. These gains show that the teachers took their work seriously and that learning did take place. The generalization can be supported by the quality of work submitted to the correspondence center. The scores on the midterm and the final examination also indicated that creditable course work had been done. Group viewing by participating teachers and the study sessions which followed brought a new dimension into the El. Ed. 426 TV, that of small-group interaction in learning about mathematics and how to teach it.

Classroom Teaching Behavior: Six of the eight OScAR (EM) scales failed to show any noticeable difference between the treatment groups. Scale 6 showed the control group made a significantly higher mean score than the experimental group. Scale 1 showed that the mean scores of experimental teachers were higher, but not significantly so. Scale 1 contained items which might be observed when the teacher was using a desirable teaching sequence. Scale 6 was made up of items which, when used sequentially, would seem to be undesirable. Since the experimental teachers used the items in

Scale 1 more often and the control teacher used those on Scale 6 more often, it is possible that some effect of the television instruction had begun to appear within the experimental group.

There may be two explanations for the similarity of scores on the other two scales. (1) The scales might not have been sensitive enough to detect simple changes in teaching behavior. Buck's (1967) study established the stability of the scales, their independence from each other, and their reliability as measuring instruments. He worked only with teachers who were not involved in any kind of instructional treatment and prepared the scales from scores made by these teachers. Items which might have shown more creative teaching behaviors possibly were not included in the scales. (2) The experimental teachers perhaps did not make substantial changes in their teaching strategies. The unofficial report of the observers (See page 35) indicated that the percentage of teaching time spent in teacher-pupil interaction was low, both for experimental and control teachers. There was no feedback from the observers, and teachers could not know what kinds of behavior were expected of them. A fifteen-week period is a rather short time to expect much change in teaching strategies. It is not easy for teachers to abandon habits of long standing in such a short period of time. Withall (1956) reported a study in which he had used deliberate feedback techniques, even to the point of getting teachers to agree to change their behavior. Later observations, however, did not show that such changes had occurred.

Conclusions: Certain teaching behaviors employed by intermediate grade teachers may undergo changes during the period of time when these teachers are taking a television course in teaching mathematics. Those changes most likely to occur consist of sequences of teaching strategies used by the teacher in conducting the class. Other behaviors investigated either did not change or the instrument used in the observations was not sensitive enough to measure the differences.

The teachers taking part in the television course made significantly greater gains on mathematics achievement test scores than those not taking the course. This fact is evidence that in-service teachers can and will work to improve their own knowledge of mathematics.

The observation period might have been too short for the observers to detect changes in teachers' behaviors. According to unofficial observation reports, both experimental and control teachers used a disproportionate part of the class time on non-interaction learning situations. Since the scales of the OScAR (EM) were designed to measure teacher-pupil interaction, the number of items which the observers might have checked was undoubtedly affected by the time spent in verbal interaction. The inertia of teachers, the tendency to follow day-to-day routines, must be recognized as a limitation of this study. It is possible that other changes did take place; some of them occurred after the visits had ended.

The R_{xy}'s between the threshold visits and the viewing period visits were low (Table 2) when scores on the eight scales were correlated. A low correlation between the pre- and post- administrations of an instrument is an indication of two possible situations: (1) The instrument is not sensitive enough (does not contain enough items) to measure all of the behaviors which occur. (2) There is little or no relationship between the measurable items which did occur on the two sets of visits. Of the two situations, the former was more likely to occur, but because different mathematical topics were being taught by the teachers under observation, as lessons followed each other, the second situation might also have been true.

Implications: Boyd and DeVault (1966) expressed a note of caution about the use of observation techniques for the evaluation of teacher behavior: (1) There is a danger of inferential analysis of observable overt behavior. (2) Consideration of only overt behavior removes the element of purposiveness. (3) Once behavior is identified, the researcher faces the problem of measurement. (4) Observed behaviors may not represent the total behavior in the classroom. The four limitations outlined by Boyd and DeVault represented most of the concerns felt by the investigator about this study. The decision to conduct this research, however, was supported by three beliefs: (1) The real measure of a teacher's effectiveness can be found only through systematic observation of that teacher's work in the classroom. (2) Observations of teachers at work are ineffective without some instrument for recording, analyzing, and

reporting their behaviors. (3) Repeated investigations, each of which builds upon another one, provide the best method of refining and perfecting instruments for studying a teacher's behavior.

Educational television stations which serve remote areas should explore the possibility of braodcasting mathematics methods courses to teachers. Teachers taking the course appear to improve their knowledge of mathematics and possibly to change some of their teaching behaviors.

SUMMARY

The purpose of this investigation was to determine whether changes in the mathematics teaching behavior of intermediate grade teachers reported while they were taking a 15-week course in teaching mathematics differed from the changes made by other teachers who were not taking the course.

WPSX-TV, the educational television station of the Pennsylvania State University broadcast an in-service course for teachers of elementary grades, El. Ed. 426, Teaching Modern Mathematics during the spring semester, 1966. The course was offered for graduate and undergraduate credit and was administered by the Correspondence Department of the University. The investigator was the instructor of the course.

Sixteen pairs of teachers from six school districts in the broadcast area took part in the study. The teachers in each pair were teaching in the same attendance unit. One teacher from each pair was chosen by lot to take the course; the other teacher agreed to take no part in any of the activities related to the course.

The design of the study was as follows:

1. To measure the mathematics achievement of the teachers in the sixteen pairs before the course began and after it was completed.

2. To make threshold visits to each of the 32 mathematics classrooms taught by these teachers to observe their teaching behaviors prior to the period of instruction.

3. To continue the visits during the viewing period, both to the classroom of teachers taking the course (experimental) and those not taking it (control).

4. To quantify the behaviors observed during these visits and study changes in teacher strategies which occurred during the television viewing period.

As a part of his doctoral research, Buck (1967) prepared an Observation Schedule and Record for Elementary Mathematics which he called the OScAR (EM). This instrument was a modification of the OScAR, which had been developed by Medley and Mitzel (1963) to study the classroom behaviors of teachers. Buck divided this instrument into eight scales, each designed to measure a different classification of teacher behavior.

Seven trained observers made 30 visits to each of the 32 teachers in the study. Seven of the visits were threshold visits and 23 of them were made after the lessons began. The observers used 20 minutes during each visit to record on the OScAR (EM) the mathematics teaching behaviors of the teacher being visited. The investigator then used the eight scales of the instrument to classify the information secured during the classroom visits.

Conclusions: Data developed through analysis of the observation reports indicated that certain changes in teaching strategies probably occurred among the television viewing teachers. The changes seemed to appear most often in the teaching sequences, or the strategies used in conducting the classes. Two of the scales which

measured these behaviors showed observable differences between the groups. No observable differences appeared on any of the other six scales.

The experimental teachers significantly outgained the control teachers on two tests of mathematics achievement, a Test of Arithmetic Principles and a Test of Mathematics Vocabulary.

Implications: A fifteen-week period was probably too short to gather complete information about behavior changes. This was a learning period for teachers, and by the time it ended, there was too little time for teachers to take full advantage of what they had learned.

There was no feedback, and teachers could only surmise the teaching behaviors the observers might have been seeking. Study of methods together with feedback from supervisors might have been more productive of change.

REFERENCE LIST

1. Barr, Avril S. Characteristic Differences in the Teaching Performance of Good and Poor Teachers in the Social Studies. Bloomington: Public School Publishing Company, 1929.

2. Boyd, R. D. and DeVault, M. Vere. "The Observation and Recording of Behavior," Review of Educational Research, XXXVI, No. 5. December, 1966, 529-551.

3. Buck, Crayton L. Mathematics Teaching Behavior of Selected Elementary School Teachers. University Park, Pennsylvania: Unpublished doctor's dissertation, The Pennsylvania State University, 1967.

4. Corle, C. G. Teaching Mathematics in the Elementary School. New York: The Ronald Press Company, 1964.

5. Corle, C. G. Test of Mathematics Vocabulary, Unpublished test. University Park, Pennsylvania, 1966.

6. Cooley, William W., and Lownes, Paul R. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley and Sons, 1962.

7. Cornell, F. G., Lundvall, C. M., and Saupe, J. L. An Exploratory Measurement of Individualities of Schools and Classrooms. Champaign-Urbana: University of Illinois, Bureau of Educational Research, 1952.

8. Deans, Edwina. Elementary School Mathematics, New Directions. Washington: U. S. Department of Health, Education, and Welfare, 1963, 116 pages.

9. DeVault, M. Vere, Houston, W. Robert, and Boyd, Claude C. Television and Consultant Services as Methods of In-Service Education for Elementary Teachers of Mathematics. Bureau of Laboratory Schools, Publication No. 15. Austin: The University of Texas, 1962.

10. Follis, Lee. The Use of Closed-Circuit Television to Improve Teacher Effectiveness. Fontana, California: Fontana Unified School District (No date given).

11. Frazier, A. and Evans, L. D. Testing the Effectiveness of Two-Purpose Television Programs in Contributing to Both Teacher and Pupil Learning (Elementary Science). Columbus, Ohio: Center for School Experimentation, University Schools, The Ohio State University, 1960.

12. Garry, Ralph, and Mauriello, Edna. Summary of Research on 'Parlons Francais': Year Two. Boston, Massachusetts: Modern Language Project of the Massachusetts Council for Public Schools, Inc., 1960.

13. Glasgow, M. W. A Study of the Relative Effectiveness of Selected Approaches to the In-Service Education of Teachers in the Utilization of In-School Radio and Television Broadcasts. Norman, Oklahoma: Extension Division, University of Oklahoma. (Not dated).

14. Hunt, Lyman C. An Experimental Project Appraising the Effectiveness of a Program Series on Reading Instruction Using Open-Circuit Television. The Reading Center of the College of Education, The Pennsylvania State University, 1961.

15. Horn, F. Distribution of Opportunity for Participation Among the Various Pupils in Classroom Recitations. Teachers College Contributions to Education, 1914, No. 67.

16. Hughes, Marie. Development of the Means for Assessment of Quality of Teaching in Elementary Schools. Salt Lake City: University of Utah Press, 1959.

17. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton-Mifflin Company, 1953.

18. Medley, D. M. and Mitzel, H. E. "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching, Edited by N. L. Gage. Chicago: Rand McNally and Company, 1963, 247-328.

19. Milgram, Joel. "Observers Report on Use of Elementary Mathematics Teaching Time," Unpublished research report: The Pennsylvania State University, Department of Elementary Education, 1967.

20. Schueler, Herbert, Gold, Milton J., and Mitzel, Harold E. Phase I: Improvement of Student Teaching. New York: City University of New York, Office of Education, Grant No. 730035, 1962.

21. Shipp, Donald, and Deer, George. "The Use of Class Time in Arithmetic," The Arithmetic Teacher, March, 1960, 117-131.

22. Shuster, Albert, and Pigge, Fred L. "Retention Efficiency of Meaningful Teaching," The Arithmetic Teacher, January, 1965, 24-31.

23. Soar, Robert S. Multivariate Statistical Procedures in Predicting Teacher-Pupil Classroom Behavior. Columbia: University of South Carolina. Cooperative Research Project No. 1170, 1962.

24. Spiker, Charles. One-Factor Repeated Measurements ANOVA, Type I. Institute of Child Behavior, University of Iowa.

25. Stoneking, Lewis W., and Welch, Ronald C. Teachers' and Students' Understanding of Arithmetic. Bloomington: University of Indiana, 1961.

26. Thomas, Dorothy S., and associates. Some New Techniques for Studying Social Behavior. Child Development Monograph, 1929.

27. Wilk, R. E., and Edson, W. H. A Study of the Relationship between Observed Classroom Behaviors of Elementary Student Teachers. Predictors of those Behaviors and Ratings by Supervisors. Minneapolis: University of Minnesota, Cooperative Research Project No. 473, 1963.

28. Withall, John. "The Development of a Technique for the Measurement of Socio-Emotional Climate in Classrooms," Journal of Experimental Education, 1949, 347-361.

29. Withall, John. "An Objective Measurement of a Teacher's Classroom Interaction," Journal of Educational Psychology, 1956, 47:203-212.

30. Wittich, Walter A. Off-Camera: A Pictorial Glimpse of AV-TV. Madison, Wisconsin: University of Wisconsin, School of Education, June, 1961.

31. Wrightstone, J. W. "Measuring Teacher Conduct of Class Discussion," Elementary School Journal, 1934, 34:454-460.

32. Wright, Muriel J. "Development of an Instrument for Studying Verbal Behaviors in a Secondary School Mathematics Classroom," Journal of Experimental Education, December, 1959, 28:103-121.

33. Zahn, Karl G. "Use of Class Time in Eighth Grade Arithmetic," The Arithmetic Teacher, February, 1966, 113-120.

Appendix A
Items of the OScAR (EM)

Appendix A

OScAR (EM)

Part I

		1.	VCBLRY, OBJCTV MTRLS		
1.	A. NMRTN	1.	B. COMPTN	1.	C. GMTRY
	bs		add		arc
	crdnl		adnd		cngrnt
	cntg,ntrl,nmbr		arry		ln
	dcml		asctv prpty		ln sgmnt
	dgt		bny oprtn		pnt
	expnd ntatn		cmn dnmtr		ry
	expnt,expntl ntatn		cmttv prpty		smpl clsd crv
	intgr		cmpstn		smpl clsd srfc
	nmbr		dfrnce		D. STS
	nm f nmbr		dstrbtv prpty		dsjnt
	ngtv nmbr		dvde		elmnt
	nmrl		enumrt		empt, nll
	ordr		eqtn		eqvlnt
	ordrd pr		eqvlnt frctn		idntl
	ordnl		fctrs		intsctn
	plc		grtr thn		st
	plc hldr		idnty elmnt		sbst
	pstv nmbr		ls thn		unn
	pwr		msrmnt		qotnt
	prpty		mltple		rto
	rtnl nmbr		mltply, mltplctn		rmndr
	rl nmbr		mr thn		rcprl, invrs
	sgn nmbr		nmbr,mthmtcl sntn		rgpng, rnmng
	varbl		oprtn		sbtrt
	whl nmbr		partn		sum
	zr		prm,prm nmbr		unknown
			prdct		X + y (is) z; x - y (is) z

I. VOCABULARY AND OBJECTIVE MATERIALS

I. A. Numeration

Base
cardinal
counting number,
 natural number
decimal
digit
expanded notation
exponent or
 exponential notation
integer
number
name for number
negative number
numeral
order
ordered pair
ordinal
place
placeholder
positive number
power
property
real number
rational number
sign number
variable
whole number
zero

I. B. Computation

add
addend
array
associative property
binary operation
common denominator
commutative property
compensation
difference
distributive property
divide
enumerate
equation
equivalent fraction
factors
greater than
identity element
less than
measurement
multiple
multiply or multiplication
more than
number or mathematical
 sentence
operation
partition
prime or prime number
product

I. C. Geometry

arc
congruent
line
line segment
point
ray
simple closed surve
simple closed surface

I. D. Sets

disjoint
element
empty or null
equivalent
identical
intersection
set
subset
union
quotient
ratio
remainder
reciprocal or inverse
regrouping, renaming
subtract
sum
unknown
x + y (is) z; x - y (is) z

OScAR (EM)

Part I
(Continued)

	VCBLRY, OBJCTV MTRLS (continued)		
1.	E. MNPLTV DVCS	1.	F. DMNSTRTV DVCS
	abcs,cntg frm		bltn brd
	clcltr		calndr
	chrnmtr		chldrn
	clck		chlkbrd-cmpts
	comps		chlkbrd-ilstrts
	cntng dvc		chrt
	csnre rds		envrnmntl obj
	dmns		flm
	flnl brd		flmstrp
	flsh crds		grph
	frctnl prts		mps
	gmtrc fgrs		mtrc chrt
	mgntc brd		mobl
	mcrmtr		nmbr ln
	ply or rl mny		obrhd prjtr
	prtctr		pstr
	rlr		thrmtr
	scl		wthr instrmnt
	sldrl		
	yrdstck		
	TCHR		Total 1
	OBSVR		Total 11
	Date		Total 111
	Time		Total IV
	Place		Grand Total

E. Manipulative Devices

Abacus or counting frame
calculator
chronometer
clock
compass
counting device
cuisenaire rods
dominoes
flannel board
flash cards
fractional parts
geometric figures
magnetic board
micrometer
play or real money
protractor
ruler
scales
sliderule
yardstick

I. F. Demonstrative Devices

bulletin board
calendar
children
chalkboard computes
chalkboard illustrates
chart
environmental object
film
filmstrip
graph
maps
metric chart
mobile
number line
overhead projector
poster
thermometer
weather instrument

OScAR (EM)

Part II

	II. Ppl Intatd INTRCTN		
II.	G. Ppl-Ppl Intrctn	II.	H. Ppl-Tchr Intrctn
	G1 agrs w clsmte		H1 agrs w tchr
	G2 dsagrs w clsmte		H2 ascts nw skl w id anthr cntxt
	G3 dfnds clsmte pnt vw		H3 exprss dsgrmnt w tchr
	G4 dfnds own pnt vw		H4 ignrs tchr qustn
	G5 cmptes on chlkbrd		H5 ofrs hypths
	G6 dmnstrts on chlkbrd		H6 ofrs posbl sltn
	G7 gvs vrbl as tnc anthr ppl		H7 rqsts mr infrmtn
	G8 rqsts anthr ppl rpt stmnt		H8 rqusts tchr rpt stmnt
	G9 verbly exprss insght, undstng		H9 sks tchr hlp
	G10 volntrly ilstrts any mdia		H10 volntr invstgt unslvd prblm
			H11 volntr gt mr infrmtn
			H12 volntr reltd, prtnt infrmtn

II. PUPIL INITIATED INTERACTION

II. G. Pupil-Pupil Interaction

G1 Agrees with classmate
G2 Disagrees with another pupil
G3 Defends classmate's point of view
G4 Defends own point of view
G5 Demonstrates on chalkboard
G6 Computes on chalkboard
G7 Gives verbal assistance to another pupil
G8 Requests another pupil to repeat statement
G9 Verbally expresses insight or understanding
G10 Voluntarily illustrates by any media

II. H. Pupil-Teacher Interaction

H1 Agrees with teacher
H2 Associates new skill with idea in another context
H3 Expresses disagreement with teacher
H4 Ignores teacher's question
H5 Offers hypothesis
H6 Offers possible solution
H7 Requests more information
H8 Requests teacher to repeat statement
H9 Seeks teacher's help
H10 Volunteers to investigate an unsolved problem
H11 Volunteers to get more information
H12 Volunteers related or pertinent information

Part III

III.	TCHR-INTATD INTRCTN	III.	K TCHR RCPTVTY
	J. TCHR-PPL INTRCTN		K1 acpts sol.tn w/o cmnt
	J1 aks ppl ilstrt answr		K2 acpts stmnt w/o cmnt
	J2 aks ppl ho h slvd prblm		K3 acptnt n-vrbl aknldgmnt
	J3 discsn mdrtr		K4 alws anthr ppl evite ans.
	J4 pos idng, strctg qustn, stmnt		K5 alws any ppl rstte cntrb
	J5 lcts ky cntrbtr, cls on hm		K6 alws ppl dcde chce sltn
	J6 mks rprvng rmrk		K7 antpts sgnfcnt dvlpmnt
	J7 rcgnzs ar ppl inadqcy		K8 dmnstrts afctn fr ppl
	J8 rfrs ppl qustn to anthr ppl		K9 encrgs altrnt sltn
	J9 rphrss, rstrctrs ppl cntrbtn		K10 ignrs ppl stmnt, qustn
	J10 rqst ppl rpt stmnt, cntrbtn		K11 lcks smpthy w ppl flre
	J11 rqst ppl evlte cntrbtn		K12 uses pstv renfrcmnt
	J12 uses srcsm		K13 wrts fr ppl frmlt thnk ans

III. TEACHER INITIATED INTERACTION

III. J. Teacher-Pupil Interaction

J1 Asks pupil to illustrate answer
J2 Asks pupil how he solved problem
J3 Discussion moderator
J4 Poses leading or structuring question
J5 Locates key contributor and calls on him
J6 Recognizes area of pupil inadequacy
J7 Refers pupil question to another pupil
J8 Refers pupil statement to another pupil
J9 Rephrases or restructures pupil contribution
J10 Requests pupil to repeat statement or contribution
J11 Requests pupil to evaluate contribution
J12 Uses sarcasm

III. K. Teacher Receptivity

K1 Accepts solution without comment
K2 Accepts statement without comment
K3 Acceptant non-verbal acknowledgment
K4 Allows another pupil to evaluate answer
K5 Allows any pupil to restate contribution
K6 Allows pupil to decide choice of solution
K7 Anticipates significant development
K8 Demonstrates affection for pupil
K9 Encourages alternate solution
K10 Ignores pupil statement or question
K11 Lacks sympathy with pupil failure
K12 Uses positive reinforcement
K13 Waits for pupil to formulate and think answer

OScAR (EM)

Part IV

IV.	L. TCHR DXTRTY W MTHMTCL IDS
	L1 ads enrchmnt for ppl
	L2 alws cls qustn, stmnt prblm bk
	L3 apls lgc mthmtcl ids
	L4 dvlps frmla, rle fr prcde
	L5 intrls mthmtcl ids
	L6 intrdcs cncpt thru prblm
	L7 intrdcs hstrcl infrmtn
	L8 mtvts thru pzl, odty, etc.
	L9 qustns crct answr
	L10 rfrs avalble enrchmnt
	L11 uses ilstrtn
	L12 uses mthmtcl rl or prncple
	L13 uses opn end prblm
	L14 uses prblm w varbles
	L15 uses ppl plng
	L16 uses scl aplctn
	Type of Lesson
	Introductory
	Ongoing, Devel.
	Drill
	Review
	Correcting Assignments
	Culminating

IV. TEACHER DEXTERITY WITH MATHEMATICAL IDEAS

L1 Adds enrichment for pupil
L2 Allows class to question statement, problem in book
L3 Applies logic to mathematical ideas
L4 Develops formula or rule for procedure
L5 Interrelates mathematical ideas
L6 Introduces concept through problem
L7 Introduces historical information
L8 Motivates through puzzle, oddity, etc.
L9 Questions correct answer
L10 Refers to available enrichment
L11 Uses illustration
L12 Uses mathematical rule or principle
L13 Uses open end problem
L14 Uses problem with variable
L15 Uses pupil planning
L16 Uses social application

Appendix B
Statistical Study of the OScAR (EM)

Buck (1967) served as research assistant to the project director of this study. As a part of his own doctoral research, he helped select items for the observation schedule and record tested the instrument, trained the observers, and supervised the team of persons who made all of the classroom visits for both this investigation and his own. A detailed report of the development of the OScAR (EM) is included here.

First, Buck identified 180 items of behavior which appeared to be related to elementary school mathematics teaching. (See Appendix A) He separated these items into six categories as follows. For ease in scoring he combined "vocabulary" and "objective teaching aids" into one section and "teacher initiated interaction" and "teacher receptivity to pupils' ideas" into another section. The organization of the items on the check list is shown below:

Section 1	79 - Vocabulary 38 - Objective teaching aids
Section 2	22 - Pupil initiated interaction
Section 3	12 - Teacher initiated interaction 13 - Teacher receptivity to pupils' ideas
Section 4	16 - Teacher dexterity with mathematics

Buck then selected two teams of two observers each to test the new instrument in several classrooms for observer agreement, structural weaknesses, and the general effectiveness of the instrument. He sent copies to a jury of fifteen authorities, nine of whom replied to the inquiry. With near unanimity the respondents agreed that the behaviors contained in the OScAR (EM) represented desirable modern mathematics teaching practices for the elementary grades.

A general resume of the specialists' comments revealed that (1) the instrument, if properly used, would differentiate between various types of teaching behavior, (2) it would have diagnostic value, and (3) it would establish a model for evaluating a lesson in mathematics. Some of the specialists observed that the instrument failed to identify teacher behaviors which caused pupils to assume responsibility for their own learning, that the instrument worked most effectively in a teacher-directed class, and that there was too little attention to the structure of mathematics and the interrelation of mathematical ideas. As events turned out, however, the classes conducted by the teachers in Buck's study were essentially teacher-dominated, and inclusion of more pupil-centered behaviors would probably not have changed the final results of the study.

After minor revisions were completed, the OScAR (EM) was duplicated and Buck began to prepare the observers to use it. The training period consisted of four three-hour sessions. The items on the instrument had been coded for security reasons, so it was necessary for the users to learn the meanings of the coded symbols. In the first training session the general use and interpretation of the new instrument was considered. The observers then listened for 20 minutes to tapes of an actual mathematics lessons; in each of four five-minute intervals they checked items on one of the four sections of the OScAR (EM). The observers then compared their results, discussed their choices, and replayed the tapes. The

observers reached approximately 85 per cent agreement on their independent observations during the training sessions.

The OScAR (EM) had been divided into four sections. The observers considered only one section at a time, giving their full attention for five minutes to the items in that section. If a behavior occurred which was listed in another section of the instrument that behavior was disregarded unless it occurred again during the time the appropriate section was in use. An item was scored only once during a visit, no matter how often it was repeated during the lesson. The observers rotated the four sections, beginning with a different one each time they visited a teacher.

Buck then used the OScAR (EM) when he visited each of 28 teachers ten times. These teachers were classified in two ways, by experience and by scores on mathematics achievement tests. Fourteen of the teachers had taught 15 or more years and fourteen had taught less than five years. Seven teachers in each experience classification made scores in the upper third of the distribution of achievement test scores and seven made scores in the lower third. None of these teachers were enrolled in the television course.

Sixteen of the 180 items were not checked at all during the 280 visits in the preliminary study. Some of the items were checked infrequently while others occurred during almost every visit. If an item appeared on less than 7 per cent or more than 90 per cent of the observation records, that item was not used. This decision was based on the assumption that behaviors which occur too often or too

seldom will contribute little toward a study of differences between teachers. After eliminating both the infrequent and the too-frequent items, 43 items were left for further study.

Since the observer either tallied an item or did not tally it, each of the 43 behaviors could be represented by the dichotomy of 0 or 1. The relationship of dichotomous variables is found by computing phi coefficients and estimating the coefficient of correlation from a tetrachoric correlation table. The formula for computing the phi coefficients was:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)\ (c+d)\ (a+c)\ (b+d)}}$$

Whenever the phi coefficient is interpreted as a coefficient of correlation, it should be recognized that such a coefficient is an underestimate of the correlation which would ensue if numerical values of each distribution were available.[9] Therefore, the phi coefficient is adjusted by utilizing a conversion table of tetrachoric correlation from the phi coefficient. This entire process was accomplished by use of the 7074 IBM computer at the Computation Center, The Pennsylvania State University. The library program, Symmetric Phi Coefficient 11.2.001 by A. Wink, was used to prepare the data.

9. James E. Wert, Charles Neidt and J. Stanley Ahmann, Statistical Methods in Educational and Psychological Research, New York: Appleton-Century-Crofts, Inc., 1954, page 302.

Cooley and Lohnes (1962) explained the theory upon which factor analysis for principal components is based.*

The examination of the phi coefficients for principal components was accomplished by use of the library program, Principal Components Analysis 11.0.003, by J. Cooley, D. Laird, and L. Pryor, Computation Center, The Pennsylvania State University. This program solves for the most dominant factors and ranks them according to the variance accounted for. In this manner, eight factors were identified for use in the study.

Since the factors which would differentiate the variance among teachers were not readily discernible, they were analyzed further by the varimax rotation. "The emphasis in varimax is on cleaning up factors rather than variables. For each factor, varimax rotation tends to yield high loadings for a few variables." The library program, Varimax Rotation 11.0.004 by J. Cooley and D. Thompson, Computation Center, The Pennsylvania State University was

---

* It was necessary to find among the items considered usable under the above selection criteria those which seemed to cluster together. If the investigator could identify items which were closely related to each other, he could use them to scale the observation instrument. To determine which, if any, items appeared to cling together the coefficients (from phi coefficient analysis) were subjected to factor analysis for their principal components. "Principal components analysis is a generally useful procedure whenever the task is to determine the minimum number of independent dimensions needed to account for most of the variance in the original set of variables." "Thus principal-components analysis not only reveals how several measures of a domain can be combined to produce maximum discrimination among individuals along a single dimension, but often reveals that several independent dimensions are required to define adequately the domain under investigation."

used. This program performs an orthogonal rotation resulting in a matrix of factor loadings. Thus, factor analysis by varimax rotation brought the factors into sharper focus.

Tables 5 through 12 show the scales developed from the eight factors with the items most closely identified with each. Since all of the 43 items which were observed often enough to be used in the data showed their highest loading in one of eight factors, Buck assigned each of the 43 items to the scale corresponding to the factor in which the highest loading occurred. He thus built eight scales for classifying observed teacher behavior. The name assigned each of the eight scales seemed most descriptive of the cluster of items contained in that Scale.

It must be pointed out here that the scales which Buck identified through factor analysis were derived solely from the 280 visits made to the 28 teachers in his study. Most of these teachers were using methods which were acceptable in the geographical and cultural region in which the study was conducted. Buck made no effort to make a qualitative evaluation of any teacher's behavior; he simply recorded their activities and reported the frequency of occurrence. If this instrument should be used in another community, especially one with a different cultural background, the scales developed by Buck should probably be reexamined before they are used.

It would be difficult to use any of these scales to show that one set of teachers performed better or worse than another.

One might perhaps generalize from the list of items on any scale that a teacher following such practices would be using desirable teaching techniques. Similarly, empirical judgment might indicate that the set of items on another scale are typical of negative or undesirable mathematics teaching behaviors. It was not the purpose of Buck's study, however, to speculate on whether the performance of a set of teachers is good or bad; rather its purpose is to show changes in teaching strategies which followed one of the two kinds of treatment identified in the experimental plan.

Tables 5 - 12 show Buck's eight scales of the OScAR (EM) and the items identified with each.

Table 5
OScAR (EM) Items Identified With Scale 1
Teacher Directing Strategies and Pupil Response

Items	Factor Loading
1. Pupil agrees with teacher.	.23
2. Pupil offers hypothesis.	.45
3. Teacher poses leading or structuring questions.	.57
4. Teacher rephrases or restructures pupil contribution.	.55
5. Teacher requests pupil to repeat contribution.	.44
6. Teacher applies logic to mathematical ideas.	.56
7. Teacher introduces concept through problem.	.32
8. Teacher uses illustration.	.63
9. Teacher uses social application.	.40
Eigenvalue	2.866

Table 6
OScAR (EM) Items Identified With Scale 2
Pupil-Teacher Interaction

Items	Factor Loading
1. Pupil gives verbal assistance to another pupil.	-.68
2. Pupil expresses disagreement with teacher.	-.56
3. Teacher allows pupil choice of solution.	-.25
4. Teacher interrelates mathematical ideas.	-.42
Eigenvalue	1.990

Table 7
OScAR (EM) Items Identified With Scale 3
Teacher Vocabulary Emphasis

Items	Factor Loading
1. Decimal	-.19
2. Number	-.34
3. Numeral	-.61
4. Zero	-.52
5. Divide	-.31
6. Regrouping	-.42
7. Teacher uses Blackboard	-.51
Eigenvalue	1.868

Table 8
OScAR (EM) Items Identified With Scale 4
Classroom Discussion Strategies

Items	Factor Loading
1. Pupil offers possible solution.	-.27
2. Pupil volunteers pertinent information.	-.35
3. Teacher asks pupil to illustrate answer.	-.44
4. Teacher locates key contributor and calls on him.	-.30
5. Teacher waits for pupil to formulate answer.	-.47
6. Teacher uses mathematical rule or principle.	-.51
Eigenvalue	1.722

Table 9
OScAR (EM) Items Identified With Scale 5
Teacher Motivating Behavior

Items	Factor Loading
1. Sum	-.51
2. Teacher uses environmental object.	-.29
3. Teacher encourages alternate solution.	-.36
4. Teacher adds enrichment for pupils.	-.42
5. Teacher develops rule or formula for procedure.	-.57
6. Teacher questions correct answer.	-.31
Eigenvalue	1.601

Table 10
OScAR (EM) Items Identified With Scale 6
Classroom Interaction in Problem Solving

Items	Factor Loading
1. Pupil computes on chalkboard.	-.30
2. Teacher asks pupil how he solved problem.	-.26
3. Teacher refers statement or question to another pupil.	-.32
4. Teacher makes reproving remark.	-.20
5. Teacher accepts solution without comment.	-.16
Eigenvalue	1.530

Table 11
OScAR (EM) Items Identified With Scale 7
Teacher Vocabulary Reference to Fundamental Operations

Items	Factor Loading
1. Add	-.69
2. Multiply	-.35
3. Remainder	-.35
4. Subtract	-.43
Eigenvalue	1.483

Table 12
OScAR (EM) Items Identified With Scale 8
Information Seeking by Pupils

Items	Communalities	Factor Loading
1. Pupil requests more information.	.25	- .39
2. Pupil seeks teacher's help.	.39	- .60
	Eigenvalue	1.466

Stability of the Scales. After identifying the eight scales through varimax rotation as described above, Buck examined each of them for estimates of stability. These estimates of stability were computed by analysis of variance using a one-factor repeated measurement (Type I Lindquist)[12] design. This is a design of one within-subject dimension (dimension A) and one between-subject dimension (dimension B). In such a design "b" repetitions represent the total number of treatment groups in the B dimension. The model for examining the data may be viewed as B representing teachers, A items and visits as "b" repetitions. Visits are nested under teachers with teachers and visits considered random and items fixed. Lindquist[13] pointed out that the Type I design can be applied when observations are

12. E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education, (Boston: Houghton Mifflin Company, 1953), page 267.
13. Ibid., page 273.

taken at regular or stated intervals and the intervals are the same. The machine programs for One-factor Repeated Measurements ANOVA (Type I) was written by Charles Spiker[14] and adapted for use on the 7074 IBM computer by Francis J. DiVesta.

A schematic model is shown in Appendix E. For each OScAR (EM) scale the model varies according to the number of items of behavior.

The estimated reliability coefficients represent the correlation between scores based on observations made by different observers at different times. Table 13 contains the reliability coefficients for each of the eight OScAR (EM) Scales.

Table 13
Reliability Coefficients for OScAR (EM) Scales

Scales	Pxx
1. Teacher Directing Strategies and Pupil Response	.74
2. Pupil-Teacher Interaction	.70
3. Teacher Vocabulary Emphasis	.71
4. Classroom Discussion Strategies	.70
5. Teacher Motivating Behavior	.72
6. Classroom Interaction in Problem Solving	.69
7. Teacher Vocabulary Related to Fundamental Operations	.90
8. Information Seeking by Pupils	.68

14. Charles Spiker, One-factor Repeated Measurements ANOVA (Type I), Director, Institute of Child Behavior, University of Iowa.

The reliability coefficient was defined as $r = \frac{\sigma t^2}{\sigma x^2}$ where tx^2 represented the variance of true scores around the mean. σx^2 represented the variance of the obtained scores of all teachers in the population around their mean.[15] In estimating the reliability coefficients with a nesting classification r_{xx} was defined as a true score of the nested parameter divided by the true score plus the error term. This procedure estimated the reliability coefficient for each scale. The reliability coefficients listed in Table 13 show that the scales were consistent in identifying certain elements of teacher behavior when successive visits were made to the same teacher's classrooms.

Correlations Between Scales. The correlations between the teachers' scores on the eight scales give some evidence of the independence of the scales. Table 14 shows these correlations.

15. Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching, American Educational Research Association, Rand McNally, Chicago, 1963, page 310.

Table 14
Intercorrelations Among the Eight OScAR (EM) Scales

	2PTI	3TVE	4CIS	5TMB	6CIPS	7TRFO	8LSP
1. Teacher Directing Strategies	.29	.34	.37	.29	.33	.34	.26
2. Pupil Teacher Interaction		.39	.11	.53	.19	.18	.12
3. Teacher Vocabulary Emphasis			-.16	.67	-.02	.42	.33
4. Classroom Interaction Strategies				.04	.28	.25	-.10
5. Teacher Motivating Behavior					.09	.30	.18
6. Classroom Interaction in Problem Solving						-.01	-.35
7. Teacher Reference to Fundamental Operations							.29
8. Information Seeking by Pupil							

The eight OScAR (EM) scales developed by Buck should have been independent of each other because they were selected from a principal components analysis. Buck, nevertheless, computed the correlations between the scales and reported them as shown in Table 14. With two exceptions (Scale 5 with Scale 2 and Scale 3) all correlations were low. Although Buck could not account for these exceptions, he concluded that, in general, the scales did operate independently of each other.

Discrimination Between Teachers. To say that the OScAR (EM) scales are stable dimension of teacher behavior does not infer that they discriminate between teachers. Buck made further analysis of the scales to determine whether greater than chance differences existed between scores assigned to teachers by observers visiting their classrooms. He used a single classification analysis of variance, AOVD, 11.5.009 by Richard Craig of the Computation Center, The Pennsylvania State University, for examining these differences among teachers. The resulting F. Ratios are found in Table 15.

Table 15
F Ratios for OScAR (EM) Scales

Scale	F Ratio	Probability (32 and 297 DF)
1	5.64	< .01
2	1.78	< .01
3	2.90	< .01
4	2.31	< .01
5	2.07	< .01
6	1.64	< .05
7	2.30	< .01
8	1.83	< .01

Buck interpreted the data in Table 15 to mean that the scales discriminated between teachers regardless of the time of the visit or the identity of the observer.

Appendix C
Pre- and Post Test Scores on Two Tests of Mathematics Achievement

Table 16. Pre- and Post-Test Scores on Two Tests of Mathematics Achievement by Experimental and Control Teachers

Experimental (T.V.)					Control (No T.V.)				
Teacher No.	Principles		Vocabulary		Teacher No.	Principles		Vocabulary	
	Pre Test	Post Test	Pre Test	Post Test		Pre Test	Post Test	Pre Test	Post Test
1	67	66	52	58	1	35	49	29	46
2	57	56	38	51	2	47	48	34	40
3	44	61	26	50	3	58	57	34	33
4	53	66	40	58	4	58	51	40	51
5	39	55	26	41	5	63	57	36	38
6	58	52	27	38	6	57	58	18	24
7	66	65	52	53	7	39	36	32	32
8	60	58	37	51	8	52	58	22	40
9	55	59	33	44	9	27	37	27	18
10	55	63	40	51	10	55	59	41	43
11	48	56	30	42	11	37	32	7	17
12	31	39	24	36	12	35	33	20	20
13	59	60	42	49	13	59	56	31	35
14	57	49	37	46	14	60	59	35	41
15	54	56	27	43	15	61	62	40	46
16	61	63	31	44	16	57	58	27	25

The maximum score on the Principles Test was 68; on the Vocabulary Test 60 was the highest score.