R842-09 ERIC REPORT RESUME

ED 010 176 1-30-67 56 (REV)
CLASSROOM SIMULATION--FURTHER STUDIES ON DIMENSIONS OF REALISM.
KERSH, BERT Y.
RQR60180 OREGON STATE SYS. OF HIGHER ED., MONMOUTH
NDEA-VIIA-1183
BR-5-0646
R-42
31-DEC-65 OEG-747-0000-228
EDRS PRICE MF-$0.18 HC-$3.48 87P.

*MODELS, *SIMULATION, *FILMS, SOUND FILMS, *STUDENT TEACHERS, TEACHERS, *STUDENT TEACHING, SUPERVISORS, SUPERVISION, CLASSROOMS, CLASSROOM RESEARCH, CLASSROOM TECHNIQUES, MONMOUTH, OREGON

AN INSTRUCTIONAL MODEL CALLED FOR A SIMULATED CLASSROOM SETTING IN WHICH A STUDENT TEACHER COULD PRACTICE NEW TEACHING BEHAVIORS UNDER SUPERVISION AND EXPERIENCE THE CONSEQUENCES OF HIS OWN BEHAVIOR. THE MODEL ALSO SPECIFIED THAT THE SUPERVISOR OF THE STUDENT TEACHERS BEHAVIOR HAVE THE CAPABILITY FOR CONTROLLING THE CLASSROOM EVENTS WHICH FOLLOW IMMEDIATELY UPON THE STUDENT TEACHERS ACTIONS. IF THE SUPERVISING TEACHER (E) WERE TO JUDGE THE INSTRUCTIONAL BEHAVIOR OF THE STUDENT TEACHER (T) AS BEING EFFECTIVE, E COULD REINFORCE TS BEHAVIOR BY CAUSING THE STUDENTS IN THE SIMULATED CLASSROOM TO REACT FAVORABLY TO T. IF ON THE OTHER HAND, TS BEHAVIOR WERE TO BE JUDGED INEFFECTIVE, E COULD CAUSE CAUSE THE STUDENTS TO REACT NEGATIVELY TO T. IN ACCORDANCE WITH THE ABOVE SPECIFICATIONS, AN EXPERIMENTAL LABORATORY WAS BUILT AND TECHNIQUES DEVELOPED FOR SIMULATING A VARIETY OF CLASSROOM SITUATIONS. THE SITUATIONS WERE SIMULATED THROUGH MOTION PICTURES AND PRINTED MATERIALS. EXPERIMENTS WERE CONDUCTED CONCERNING IMAGE SIZE, FEEDBACK MODE, AND RESPONSE MODE. SEVERAL FOLLOWUP STUDIES AND ONGOING RESEARCH EFFORTS WERE ALSO DISCUSSED. (JC)

Classroom Simulation:

Further Studies on Dimensions of Realism

Bert Y. Kersh

Final Report

Title VII Project Number 5-0848

National Defense Education Act of 1958

December 31, 1965

Teaching Research Division

Oregon State System of Higher Education

Monmouth, Oregon

Acknowledgments

Dr. Paul A. Twelker, Teaching Research Division, revised the orientation and rating procedures, and directed the Follow-up Studies.

Dr. Lyle E. Calvin, Oregon State University; Dr. Paul A. Twelker, and Dr. James H. Beaird, Teaching Research Division, served as statistical consultants.

Dr. Jesse H. Garrison, Oregon College of Education, gave his cooperation and assistance as instructor of the college courses from which experimental subjects were selected.

Mr. Jack H. Bond and Mr. Victor E. Lund, Teaching Research Division, were the Experimental Instructors and provided valuable assistance generally.

Mr. Fred C. Crowell and Mrs. Virginia Weigel, Teaching Research Division, assisted in the statistical computations.

Mr. Timothy J. Pettibone, Teaching Research Division, served as laboratory technician for the project.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (Cont'd.)

Table

LIST OF FIGURES

I Problem

The studies reported herein developed from an earlier effort to employ simulation techniques in the education of classroom teachers. The problem was to provide a medium of instruction from which student teachers could learn to transfer their knowledge of teaching gained from textbooks and lectures to the actual classroom situation. The increasing difficulty teacher educators are having in providing opportunities for supervised experience in classrooms is, perhaps, reason enough for looking to classroom simulation as a laboratory experience for teachers. More fundamentally, however, the problem is to instruct teachers so that they will actually practice what they are taught.

Originally, the simulation technique developed for this present project was based on the operant conditioning model. Although a different model is now employed, the original concept better explains early developments. In the original conceptualization, certain assumptions were made. First, it was assumed that student teachers previously have acquired highly differentiated ways of behaving in the classroom which very quickly become evident when the student engages in practice teaching. Consequently, their knowledge of textbook recipes for teaching subject matter and their knowledge of model teaching behavior gained through classroom observations frequently is not evident in the classroom, at least in the beginning stages of their teaching experience. More technically, it was assumed that the classroom behavior of most student teachers is controlled primarily by external stimuli in the classroom, not by their knowledge of model teaching behavior.

Secondly, it was presumed that the classroom behavior of teachers is patterned or "shaped" by stimulus events in the classroom which act as reinforcers. Accordingly, actual changes in the instructional habits of a teacher would depend on whether or not his attempts to change are reinforced by events which follow immediately upon his actions.

From this line of thinking, it was a short step to the conclusion that, by controlling the stimulus events in the classroom while a student teacher is instructing, a teacher educator can change the classroom teaching behavior of the student more effectively than by lecturing or by demonstrating. If the teacher educator were able to control the situation precisely, theoretically he could effectively "shape" the behavior of the student teacher in the classroom by reinforcing successive approximations of the desired behavior.

The original problem, therefore, was to develop an instructional medium which would meet the essential requirements of the operant conditioning model. As conceptualized above, the instructional model called for a simulated classroom setting in which a student teacher could practice new teaching behaviors under supervision and experience the consequences of his own behavior. At the same time, the instructional model specified that the supervisor of the student teacher's behavior have the capability for controlling the classroom events which follow immediately upon the student teacher's actions. If the supervising teacher (hereafter referred to as "E" for "experimenter") were to judge the instructional behavior of the student teacher (hereafter referred to as "T") as being effective, E could reinforce T's behavior by causing the students in the simulated classroom to react favorably to T. If on the other hand, T's instructional behavior were to be judged ineffective, E could cause the students to react negatively to T.

In accordance with the above specifications, an instructional laboratory was built and techniques were developed for simulating a variety of classroom situations. The situations were simulated through the medium of sound motion pictures and printed materials. Multiple projection techniques were employed to create the desired realistic effects, and to enable E to effectively control the events which followed T's instructional behavior.

As would be expected, it was not within the practical limitations of the original project effort to completely fulfill the requirements of the operant conditioning model. Instead of "reinforcing" T's behavior, E projects a motion picture showing the "most likely consequences" of T's behavior. That is to say, he is shown the way his class most likely would respond under the circumstances. This "feedback" information may or may not be reinforcing, and it may or may not have a controlling influence on T's behavior. However, E does have the ability to choose the type of information which is provided the learner. Thus, the instructional medium is said to employ "controlled feedback."

Controlled feedback now is considered simply as information which T may incorporate in an ongoing problem-solving, decision-making, or hypotheses-testing thought process. T's instructional behavior is not "shaped" in accordance with the operant conditioning model. Nevertheless, T's instructional behavior is changed in accordance with prescribed instructional objectives, and there is preliminary evidence to indicate that there are positive transfer effects to the actual classroom situation resulting from simulation instruction.

The present research effort examined selected instructional variables in the classroom simulation technique, as originally developed, so as to provide a sound basis for further development of the instructional medium. Since transfer effects of the instructional medium were of primary concern, those instructional variables contributing to the fidelity (realism) of the simulated classroom were investigated first.

By the learning model of operant conditioning, the stimulus feature of the simulated classroom reasonably should correspond closel o those of the classroom to which T is to respond differentially. Also, T should learn to behave in the simulated classroom precisely as he is to respond in the actual classroom.

On the other hand, the information systems model (Ryans, 1963), which characterizes the present classroom simulation technique, is more nearly cognitive and places fewer constraints on the physical characteristics of the instructional system. Ryans' information systems model emphasizes those dimensions of the instructional system which affect cognitive learning (knowledge of the instructional standards) and which develop skill in making decisions on the basis of feedback from pupils in the classroom. Reasonably, a student teacher could learn instructional principles and could develop skill in discriminating one form of student behavior from another from either life-size or smaller projections of the children's behavior, and would not necessarily have to act out his responses in the simulated classroom. If so, the simulation materials developed for use in the present simulation laboratory facility could be adapted for use in a more conventional classroom setting using standard projection equipment, or could be adapted for self-instruction where the student teacher would require a minimum of supervision while actually responding to the motion picture sequences.

Findings of the recently completed research on dimensions of realism, therefore, have important theoretical as well as practical implications. If the transfer effects of instruction in the simulated classroom are not related to instructional variables involving size of image and mode of response, the same instructional materials could be adapted for use on a broader scale and at lower cost.

II The Classroom Simulation Technique

Simulation as an instructional technique has been successfully employed in other walks of life, principally in the industrial and military areas. Simulation techniques are employed for practical reasons, either to avoid damage to costly equipment or to avoid endangering lives, as in the training of air defense personnel and astronauts. Through simulation techniques, commercial pilots are checked out in modern aircraft and air force officers learn to direct fighter interceptions of enemy bombers (Adams, 1962; Chapman, 1959; SAGE Training Group, 1957). They learn through actual experience in situations which approximate reality. The sensations of sight, sound, and motion are so realistic that the learner frequently has difficulty retaining his true orientation in the artificial setting. Computer-based "games" and related simulations place less emphasis on the physical setting, but are equally good examples (Guetzkow, 1962; Greenlaw, et al, 1962).

However, simulation techniques have only recently been employed in the professional field of Education. One explanation for this relatively slow adoption of a promising instructional technique by Educators may be that, until recent years, it has been relatively easy to arrange laboratory experiences for student teachers. Actually, simulation techniques were first employed in Education for research purposes. The Whitman School simulation materials, for example, now fairly commonly used as a laboratory experience for school administrators, actually were developed for purposes of evaluation. The Whitman School materials simulate an entire school setting through the use of motion pictures, tape recordings, and printed documents (Cunningham, 1959; Frederiksen, 1962; University Council for Education Administration, 1960). The participants play the role of principal in the Whitman School. After becoming thoroughly oriented to the hypothetical situation, the participants sit at a desk and solve administrative problems, one at a time, which are presented via their "in-basket." The participants respond to a great variety of problems ranging from a new teacher who asks to be released from her contract, to an emergency situation which occurs on the school grounds. They write out how they handle each situation so that their performance can be evaluated subsequently. Although not designed primarily as an instructional tool, the same materials have been employed extensively in seminars as a basis for discussion and analysis of typical school problems encountered by school administrators.

The classroom simulation technique employed in the present research was developed first as an instructional medium, but not primarily as a means for resolving practical problems encountered by teacher educators in providing laboratory experiences for student teachers. The original objective in developing a simulated classroom for training teachers was to augment conventional classroom laboratory experiences through the provision of a more systematic, controlled learning experience.

Instructional materials for classroom simulation. Classroom simulation attempts to create for the student teacher all of the relevant features of a single classroom situation. Potentially, many different classrooms could be simulated, but presently the materials are limited to one group of sixth graders named "Mr. Land's Sixth Grade." Mr. Land is the hypothetical supervising teacher with whom the student teacher works during his simulated experience. The simulation materials include a complete set of cumulative records for each of 22 youngsters in the hypothetical class. Each cumulative record file includes standardized test data, achievement records, health records, a summary of anecdotal records kept by Mr. Land, and a snapshot of the child. In addition, there are printed descriptions of a hypothetical school, called "College Grove Elementary," and a community named "College Grove." The main body of materials used in the instructional phase includes orientation films showing Mr. Land working with his class in a typical fashion, and 60 problems on film, each of which poses situations to which T reacts as if he were the actual teacher. Each problem situation on film comes with alternative "feedback sequences" designed to show T how Mr. Land's pupils might react to him under different circumstances. A single group of youngsters was used throughout the filming, and they were filmed so that they appear to be reacting directly to T as he views the problem sequences.

The 60 problem sequences are divided into three sets of 20 sequences called Program I, II, and III. Each of the three programs correspond to one school day and are roughly parallel in terms of types of problems included. Half of the film sequences pose problems in classroom management for the student teacher, and the remaining are classed as communication problems. More detailed descriptions are included in the appendix.

Instructional procedure. The procedure described below is the procedure which has been used most recently in the present research. The student teachers (Ts) are instructed individually in a special laboratory facility. The simulation facility is diagrammed in Figure 1.

First, T is oriented to the simulated classroom, "Mr. Land's Sixth Grade," and to the instructional procedure. Ts are oriented with a slide-tape description of the classroom and a self-instructional program designed to acquaint them with the children in Mr. Land's class. The orientation includes an experience in the laboratory facility during which T observes Mr. Land interacting with the children (on film) and during which T is asked to "introduce" himself to the children.

Immediately following the orientation sequence, T is given a performance test in the simulated classroom using one of the three instructional programs as stimulus materials. T reacts to each of 20 problem sequences, and a specially prepared rating procedure is employed to assign numerical values

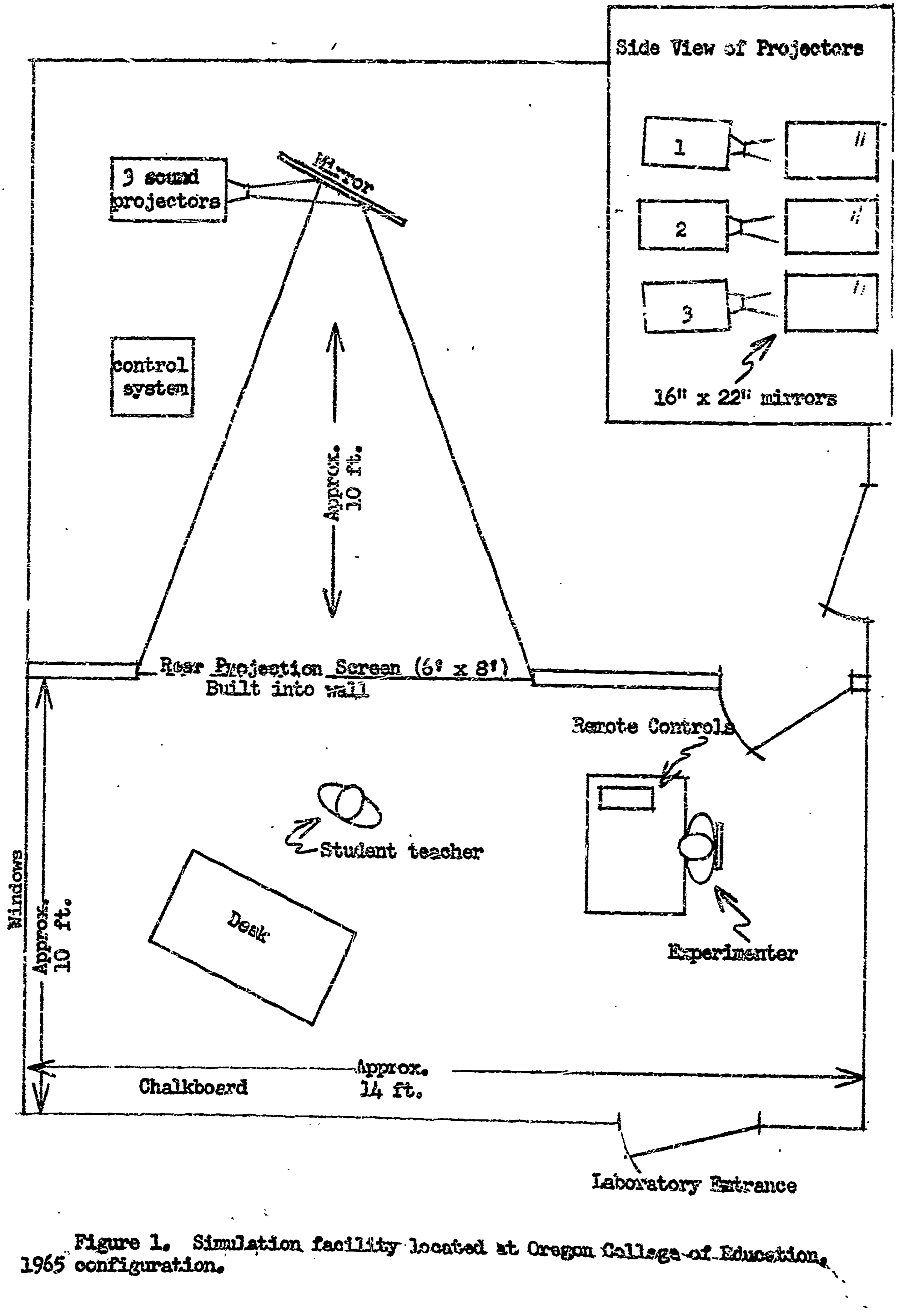

Figure 1. Simulation facility located at Oregon College of Education, 1965 configuration.

to his performance. The rating procedure is described in more detail in the appendix. Approximately one hour is required to complete the pretest.

Next, actual instruction in the simulation facility begins. A problem situation is projected on the screen, and T is requested to enact his response. Depending on T's response, one of the alternative feedback sequences is projected immediately following T's response. The feedback sequence shows the pupils reacting to T in a particular way. Immediately following the projection of the appropriate feedback sequence, E begins a line of questioning aimed at (1) determining how T perceived and diagnosed the problem, and (2) assisting T in evaluating his own performance in the light of information gained through the feedback sequence. E attempts, through his questioning, to direct T through a careful analysis of the problem and to consider alternatives to the way he responded to the problem. Alternative responses are considered whether or not his response satisfies the particular requirements employed in the rating procedure. Each time T considers an alternative response to a particular problem, he is encouraged to enact it. This is accomplished by reversing the projector and showing the problem sequence over again, this time projecting a different feedback sequence. This cycle of problem-response-feedback-discussion is repeated with each problem until both T and E are satisfied that T has gained as much as can be expected from the particular problem sequence. Then instruction is advanced to the next problem on film. Typically, five to eight problems are completed during one 50 minute period in the simulation facility. Each T is scheduled for instruction in the simulation laboratory one or two hours each week, and his instruction is continued through one program consisting of 20 problem sequences.

After the instructional phase, T's performance is tested again in the simulation facility using the last of the three sets of instructional programs. The total time for orientation, pretesting, instruction, and post-testing is usually six to eight hours, not including time spent out of class studying the cumulative record files of the children.

Purpose of feedback sequences. The purpose of the feedback sequences is to communicate to T the most likely consequence of his behavior. For example, one problem sequence shows a game of "keep away" begin spontaneously in the classroom after recess period. T may try several ways to regain order. The alternative feedback sequences show the youngsters settling down in their seats or continuing to play their game. Most problems have two alternative feedback sequences, but some have as many as three or four.

Some problem sequences do not lend themselves to the preparation of "positive" or "negative" feedback. For example, in another episode, Jack approaches T and reports that he has been sick and should not be allowed to play during recess. Usually, T either asks Jack for a note from home or responds in some way which does not require Jack to present a note. The alternative feedback sequences show Jack either presenting a note or returning to his seat. The feedback sequences in this case are simply an extension of the interaction between pupil and teacher, and provide T with a concrete example to discuss with E and to evaluate in terms of supporting records and the particular context of the problem situation.

Rating standards. Rating standards were developed initially by a jury of master teachers, in connection with the first research and development effort described elsewhere (Kersh, 1963). The original set of instructional materials and rating standards have since been revised by the project staff. The list of revised standards together with representative problem sequences from the instructional programs are included in the appendix. The standards are actually rules of procedure, applicable to problems of classroom management and communication. Each standard is stated so as to make the behavioral alternative clear. What is considered desirable behavior is contrasted with what would be considered undesirable. The first standard, for instance, covers classroom problems involving rules of procedure when T is not informed of the rules. The standard states that in problems involving rules of procedure, T should defer to a person in authority; he should not establish his own rules. It is important to remember that T is presumed to be a student teacher who is being supervised by Mr. Land, the regular classroom instructor.

While revising the original standards and instructional procedure for purposes of the present research, it became evident that most problem sequences involve more than one standard. For example the first problem sequence in Program I involves two standards. Jack's statement that he has been ill for the past week and should not be allowed to play at recess is the case in point. According to present standards, the most effective way to handle the situation is (1) to communicate to Jack that Mr. Land will take care of the situation, and (2) to respond in a manner which would be judged to be "supporting." The point is, T is provided no basis for making a decision in the matter. For all he knows, the school authorities may have already established rather definite rules regarding such matters, or Mr. Land may have already been in direct communication with Jack's parents. Consequently, the standards suggest that T simply accept Jack's message at face value and assure him that his problem will be resolved.

A rating of three (3) is assigned when T's behavior is considered effective by both standards; a rating of two (2) is assigned when one of the two standards is met but not the other; and the rating of one (1) is assigned when T's behavior is considered ineffective by both standards. A zero (0) is assigned when T fails to respond at all to the problem. Variations in the rating procedure are employed with problems which involve only one standard, or three standards. These variations are described in more detail in the appendix.

T's assessment of each problem is rated by recording his description of the stimulus situation and tallying the number of items of information which correspond to each of those listed for each problem sequence. The selection of salient items of information was made by the project staff using the standards for each problem as criteria. In addition to the list of salient items of information in the stimulus situation (the film), some problems involve information which is included in the cumulative files or which was transmitted previously in the particular simulated "day" (a program of 20 problems). For example, in Program I, Jack is the key figure in several problem sequences. In each problem after the first one involving Jack, it is considered important that T state the fact that Jack was sick during the previous week. Without this information, Jack's behavior may be misinterpreted.

A composite rating for each T is made by summing the numerical ratings assigned T. A minimum of 20 problem sequences is used for a basis for evaluating T's performance.

Testing procedure vs. instructional procedure. Generally, numerical ratings are not assigned during instruction. Instead the standards and problem assessment criteria are used as examples of terminal behavior. As indicated previously, during instruction each problem sequence is repeated until T's performance reaches the criterion for the most effective response and problem assessment. However, other alternatives to those indicated in the standards are often considered in the discussions between E and T. When T's performance is being tested, the discussion following each problem is eliminated except in so far as necessary to determine T's assessment of the problem. Also, the feedback sequences are eliminated when the materials are used in the testing mode.

Questioning technique. During instruction and testing, it has been considered important that E not reveal through his questions any information which would be of value to T in his effort to learn the most effective response and to properly assess the problem. The questioning technique is in many respects comparable to that which is used during inquiry phase in projective testing (for example, the Rorschach). E is instructed to guard against leading questions which reveal information inadvertently. Examples of questions that are considered "neutral" as contrasted with "leading" or "revealing" questions are listed below:

Neutral Questions	Leading or Revealing Questions
"What was the problem? Describe it to me."	"What did Jack say to you?"
"Can you tell me more about it?"	"Do you think the school has any rules?" for handling this kind of situation?"
"What else about the situation do you think is important?"	"How did the class react to Karen?"
	"Do you remember what happened to Jack earlier in the day?"

During instruction, it is sometimes necessary for E to prompt T when he fails to make progress in his effort to learn the desired behaviors. The particular prompts provided have not been standardized to date. As a general rule, T is instructed to rely as much as possible on existing records made available to him during instruction, and on feedback provided regarding the manner in which the children in Mr. Land's sixth grade respond to him. T is expected to try out different response methods until he "discovers" a satisfactory mode of responding.

Typically, after T has responded to a particular problem effectively according to the rating standards, E stated the standards to T. However, the instructional procedure was not explicit in this regard during the earlier experiments. Characteristically, E simply acknowledged that T's performance was in accordance with that which was recommended by the "jury of experts." Under no circumstances, however, was T informed of the rating standards prior to correctly performing in accordance with them.

Detailed instructions for each problem sequence. Detailed instructions for each problem sequence were prepared so that E could refer to them directly during instruction or testing. The format was designed so as to enable E to identify the necessary information quickly and accurately. At the top of each set of materials for a particular problem sequence, a description of the situation is provided, typed exactly as it was to be communicated to T. Next, a description of the problem scene is typed to serve as a reminder for E. Of course, the problem scene was not communicated to T before being projected. As part of the problem description, a "hold cue" is indicated which specifies where the film is to be stopped in the event that T does not respond while the problem sequence is being projected. Beneath the "hold cue" instructions, standards and rating criteria for evaluating T's response are detailed.

Examples of the detailed instructional procedures for each problem sequence are included in the appendix. All instructional procedures are written for use with a specially designed control system.

III Related Research

Related efforts to develop simulation as an instructional medium were reviewed briefly in the introductory paragraphs above. A more complete review has been written by Vlcek (1965).

Perhaps one of the most extensive developments involving simulation has been in connection with the System Training Program (STP) of the System Development Corporation (SDC). From the initial studies conducted in RAND's System Development Laboratory (Chapman, et al, 1959) it was concluded that Air Force crews learned to operate an air defense center most efficiently when they are allowed to practice the operation of the system in a realistic setting and to evaluate their own performance after each exercise.

Since the earlier studies, much has been written at SDC about the program, but many questions still remain unanswered. Evidently one of the key factors contributing to the success of STP is the provision of information to the learners about the events which follow their actions in the simulated setting. This information, called "feedback," permits the learner to evaluate his own action and to decide whether to continue doing the same thing or to try some other way of handling the situation.

A research effort completed at Oregon's Teaching Research Division (Kersh, 1963) tested variations in size and motion in the projected classroom scenes of the original "Mr. Land's Sixth Grade," (the first version of the simulated classroom used in this present research). A total of 40 college juniors were assigned randomly to four experimental groups and trained in an earlier development of the simulation facility at Oregon College of Education. The "realistic" display employed life-size sound motion pictures, and the least realistic employed still projections greatly reduced in size. Performance ratings before and after instruction were used to compare the effectiveness of the more or less realistic projections in the simulation. Other comparisons were made on the basis of the number of trials required to reach criterion during the instructional phase, and the self reports of Ts' reactions to changes and display mode and to the medium as a whole.

Analysis of variance techniques were used in the statistical analysis. The smaller, less realistic display was found to be more effective than the life-size projection in terms of the Ts' post-test scores ($p < .05$), and more practice trials were required with motion pictures than with still projection ($p < .01$). However, Ts reported that the more realistic mode of presentation produced sensations of tension, fear, and frustration which were notably absent when the less realistic mode was used.

These findings are not inconsistent with previous studies with flight simulation which indicate that high fidelity is not of particular importance (Adams, 1962). It was concluded, however, that the data from the classroom simulation study actually may have reflected a complex interaction between T's initial exposure to the more realistic display and his individualized instruction in the less realistic mode. Prior to the instructional phase of the experiment, one and one half hours were devoted to orientation and testing, all of which employed the more realistic large motion picture projection. The experimental design failed to control for the learning which may have resulted from this experience; all Ts were oriented and pretested in the same manner. The total effect may explain the superior performance of students having instruction in the small-screen mode. Further experimentation was recommended to determine whether it is necessary to provide experience with the realistic followed by the less realistic in order to produce the effects revealed by the previous experiment.

It is entirely possible that the student teacher first should experience the emotional impact of a classroom of children in a variety of stressful situations and then practice effective ways of teaching under more relaxed (although not less specific or detailed) circumstances. Perhaps it makes little difference during actual instruction whether the stimulus materials are more or less "realistic" just so they provide T with the opportunity to analyze and to enact responses repeatedly under supervision.

It is also likely that results of a comparable performance test in a real classroom would be different. T's performance in a real setting is, of course, the ultimate criterion. The transfer effects of practice in the more realistic simulation, if any, may not become evident until T is observed during actual practice teaching.

Vlcek (1965) has also studied the effects of the classroom simulation facility using "Mr. Land's Sixth Grade," as instructional materials. Vlcek used the revised films and techniques which were employed in the present research effort. Although he did not manipulate the instructional procedure so as to provide for more or less realism in the display, he did study the transfer value of simulation instruction in a real classroom setting.

Vlcek selected student teachers from a junior level elementary education course at Michigan State University and assigned them randomly to an experimental and a control group, each of which was classified in terms of high and low grade point averages. The experimental group received approximately nine hours of classroom simulation experience in a laboratory training facility modeled after the one employed in the Oregon studies conducted by the present writer. Vlcek's experimental group received nine hours of classroom simulation experience while his control group received an orientation session only. Neither group was pretested in the classroom simulation laboratory, but each was tested in the simulation laboratory after completing their respective treatments (orientation only or orientation plus instruction), and were rated by a procedure based on that which was employed in the present research. Vlcek's experimental group performed significantly better than did his control group in acting out their responses to post-test problems. However, no statistically significant difference was found in the two groups' ability to accurately describe the problem after viewing the filmed sequence.

Vlcek selected Ts who were to enroll in student teaching the semester immediately following their experience in the simulated classroom. This procedure is in contrast with those employed in the present study wherein Ts were observed during student teaching approximately one year later. Vlcek found that in the actual classroom his experimental group applied a greater number of the teaching principles (standards) taught in the simulated classroom than did the control group. Also, evidence recorded by Ts on a "confidence scale" and "attitude scale," provided strong support for Vlcek's hypothesis that classroom simulation increases Ts' self-confidence in ability to teach, and indicated that Ts consider simulation to be very valuable.

On the negative side, Vlcek's findings indicate that simulation training does not insure that Ts will be more effective in directing the classroom behavior of their pupils. Although his experimental group applied a greater number of the teaching principles taught, their efforts were also more frequently judged to be ineffective. In other words, Ts who did not have experience in the simulated classroom employed fewer "teaching principles" but also were judged to have greater success in their limited efforts.

Vlcek's findings regarding the transfer effects of simulation instruction could be interpreted to mean that student teachers do learn principles of instruction which are transferred to the classroom, but that the particular principles learned in Mr. Land's Sixth Grade are not effective. This, in fact, is the interpretation which Vlcek makes. If so, it must be concluded that the "experts" who established the standards were wrong. Another possible interpretation is that his findings indicate that teachers who try a greater number of different ways of handling problematic situations in the classroom may fail in their attempts more frequently than teachers who experiment less frequently. In any event, and in the opinion of the present writer, it is far more important to know that teaching methods learned from simulation instruction do transfer to the classroom. New standards may always be developed.

Dimensions of realism. It is not possible to study all variables which might conceivably contribute to the "realism" of Mr. Land's Sixth Grade nor would the research "yield" from an exhaustive study be likely to warrant the effort even if it were possible. For example, some variables listed by Ryans (1963) as "external information inputs" which are obviously important for some problems were purposely not included in the present simulation, because the particular classroom problems employed in "Mr. Land's Sixth Grade" did not require the information. Also, the decision was made not to study such variables as the description of the school and community because it was considered unlikely that a more detailed description would alter the learning experience, and nothing of great developmental or theoretical importance would be gained from such a test.

Other variables such as size and motion in the projection are more likely to influence learning in the simulated setting. Also, from the practical standpoint, it is of extreme practical importance in the development of simulation techniques involving visual projections to know whether or not it is necessary to project a life-size motion picture. If small images are satisfactory, the development of low-cost simulation techniques would be greatly enhanced.

Consequently, selection of the variables to include in the present develop-developmental research effort (using Mr. Land's Sixth Grade as the research vehicle) was made on the basis of both practical and research expediency.

In terms of Ryans' model of the "teacher information processing system" (Ryan, 1963), the following independent variables pertaining to realism in simulation were considered worthy of further study:

1. External information inputs:

   Life-size (realistic) or small (unrealistic) projections of classroom problems during (a) orientation and pretesting, and (b) during actual instruction.

2. Feedback inputs:

   Feedback information communicated verbally (unrealistic) or by sound motion pictures (realistic).

3. Student teacher information processing:

   Response to problems enacted by T (realistic) or verbalized (unrealistic).

IV Experiment I: Image Size and Feedback Mode

The first experiment was designed to answer questions concerning three instructional variables: (1) the size of projected image during orientation and pretest phase, (2) the size of projected image during instructional phase, and (3) feedback mode, whether communicated by motion pictures (visualized) or described by E (verbalized).

One objective was to provide evidence to corroborate the findings of the earlier experiment conducted by this present writer (Kersh, 1963), using new and improved motion picture films and refined testing procedures.

A second objective of the experiment was to determine whether the size of the motion picture projection during orientation and pretesting influences learning as measured by post-test performance. Previous findings (Kersh, 1963) were based on evidence which did not control for the possible effect of orientation-pretest image size on learning. It has been suggested that there may be an interaction effect between image size employed during orientation and pretesting, and that which is employed during instruction (Kersh, 1963, p. 51). Ts reported in previous research that the more realistic mode of presentation produced sensations of tension, fear, and frustration which were not reported when they worked with the small motion picture projections. Perhaps if small size projections were to be employed throughout the orientation and pretesting, as well as the instructional sequences, the findings of the previous experiment would not have favored the use of small projected images.

The t' ird objective of the experiment was to test the relative importance of providing "realistic" feedback to T regarding the possible consequences of his behavior in the simulated classroom. For this test, motion picture projections (visual feedback) were compared with descriptions by E (verbalized).

Actually, "visual feedback" was sometimes interspersed with "verbal feedback," (see instructions in appendix). However, motion pictures predominated the visual feedback mode.

A total of eight combinations of the instructional variables in a 2 x 2 x 2 factorial design permitted a test of both the interaction and the main effects of the three variables. The eight treatment groups were identified as follows (see Table 1):

1. <u>Large-large (vis)</u>: Large (life-size) projections during orientation and pretest; large projections employed during instruction; and visualized feedback (motion picture projections equivalent to that used during instruction).
2. <u>Large-small (vis)</u>: Large projections employed during orientation and pretest; small projections (approximately one foot by two feet); during instruction; with visualized feedback.
3. <u>Small-large (vis)</u>: Small projections employed during orientation and pretest; large projections during instruction; with visualized feedback.
4. <u>Small-small (vis)</u>: Small projections employed during orientation and pretest; small projections during instruction; with visualized feedback.
5. <u>Large-large (ver)</u>: Large projections during orientation and pretest; large projections during instruction; with verbalized feedback (described by E, without the use of motion pictures).
6. <u>Large-small (ver)</u>: Large projections employed during orientation and pretest; small projections; with verbalized feedback.
7. <u>Small-large (ver)</u>: Small projections employed during orientation and pretest; large projections during instruction; with verbalized feedback.
8. <u>Small-small (ver)</u>: Small projections employed during orientation and pretest; small projections during instruction; with verbalized feedback.

Table 1

Treatment Combinations for Experiment I

(n = 80)

	Orientation-Pretest Image Size			
	Large		Small	
Visual Feedback	10*	10	10	10
Verbal Feedback	10	10	10	10
	Large	Small	Large	Small
	Instruction Image Size			

* Number in each treatment combination group

Questions to be answered. The experiment was designed to answer specific questions, and comparisons to be made in the analysis were planned accordingly. The following questions were to be answered:

1. Do the groups which were oriented and pretested in the large projection mode differ in post-test performance from those which were oriented and pretested in the small projection mode?

2. Do the groups which received visual feedback during instruction differ in their post-test performance from those who received verbalized feedback?

3. Do groups which were instructed in the large projection mode differ in their post-test performance from those which were instructed in the small projection mode?

4. Are there any differences in post-test performance among the various treatment combination groups indicating that particular interaction effects occur when the instructional treatments are combined?

The statistical analysis consisted of translating the questions into null hypotheses, then testing the hypotheses using a probability equal to or less than .05 as criterion ($p \leq .05$). Due to practical problems in data gathering, the sample size was necessarily small, but it was presumed from previous findings (Kersh, 1963) that mean differences in the magnitude of three (performance test total scores) would be statistically significant with samples of size 20. This assumption was corroborated by calculations of the estimated power of the statistical test to be employed, using the procedures outlined by Winer (1962, p. 104).

Selection of subjects. Student teacher (Ts) were selected from a junior level course called "junior block" at Oregon College of Education. Data gathering was continued for four quarters, beginning winter quarter, 1964. Participation in the classroom simulation instruction was made a requirement of the course along with other practicum experiences such as observing in elementary classrooms, assisting on the playground, and teaching a short lesson in reading to a small group of elementary school students. The subject-matter content of the course was educational psychology and general methods of teaching at the elementary level.

Although every student in the junior block course was provided instruction in the classroom simulation facility, those who did not participate as subjects in the experiment were screened out on the following basis:

1. Those over 25 years of age were eliminated.
2. Those with any teaching experience beyond incidental experience as a high school cadet teacher or Sunday-school teacher were eliminated.
3. All those scoring 34 and above on the pretest were eliminated.

At the beginning of each quarter, Ts were screened and assigned to experimental treatment combination groups by random procedures. The following assignment procedure was employed each term:

1. All eligible Ts in the junior block course were ranked by scores on the College Entrance Exam Board Test, and separated by sex.
2. By dividing each list into levels and assigning at random from each level, Ts were assigned (a) to a large or small image pretest group, and (b) to receive one of the three sets of problems (Program I, II, III) as the pretest.
3. After pretest scores were available, those scoring 34 and below on the pretest were assigned to the four remaining treatment combination groups within each pretest group.
4. Finally, each T was assigned to one of the two experimental instructors (Es). After the entire class was scheduled for the quarter, the weekly schedule was divided into two-hour time blocks, and the two Es alternated throughout each day, taking whoever was scheduled within their particular time period. Ts did not have a choice of instructor.

Data gathering was continued until a minimum of ten Ts successfully completed instruction in each of the eight treatment combinations employed in the experiment.

Orientation. Student teachers in the junior block course were introduced to classroom simulation during the first and second weeks of each term. During a regular class period, the entire class was introduced to "Mr. Land's Sixth Grade," and were told generally how instruction would proceed during the term. The class was told that they were participating in an experiment and that slightly different instructional procedures would be employed with different individuals. At the same time, they were advised that everyone would have a profitable instructional experience, i.e., that there would be no "control groups" involving substitute types of instruction. During the first meeting with the class, the printed materials describing the school and the community, and the cumulative record files were made available on a check-out basis. Students were instructed to study the materials in preparation for their first scheduled session in the classroom simulation laboratory. Finally, Ts were scheduled for the first session in the laboratory facility.

The following week, each T was further oriented and pretested individually during a one-and-one-half hour session in the classroom simulation laboratory. During this session, each T was shown a motion picture sequence of Mr. Land and his students, was allowed to practice identifying students by name, and was asked to "introduce himself" to the class while the motion picture showed the students listening attentively to T. This final orientation sequence usually lasted approximately one-half hour, and was followed immediately by the pretest.

The pretest consisted of 20 problem sequences to which T was asked to respond by enacting how he would handle the situation. Each problem was shown only once, and discussion of the problem was limited to questions asked by E, concerning T's assessment of the particular problem. No feedback sequences were shown. The pretest of 20 problem sequences was typically completed within one hour.

Instructional Phase. Instruction was begun the week following completion of orientation and pretesting, using a different set of problems (randomly selected from one of the two remaining sets of 20 problem sequences). T was allowed to practice each problem repeatedly until his performance could be rated as "most effective" by the prescribed standards.

Instruction was started for each problem sequence by E reading from the instructional materials that information which described the "setting" of the problem. The setting included the particular activity in which the class was engaged at the beginning of the problem, the location of Mr. Land, T's position in the classroom, and any other pertinent information.

Next, E started the motion picture projection of the problem and observed T's reaction to it. At the point when T began his response, the forward motion of the projector was stopped on a single frame, making it appear to T as though the children had suddenly been "frozen."

As soon as T completed his response to the problem, E selected a feedback sequence (or described it, depending on the experimental treatment involved).

After the feedback sequence, E began a line of inquiry with T during which he asked T to assess the problem and to evaluate his own response in accordance with the feedback. Frequently, this discussion led to consideration of information contained in the cumulative record files of the children, or to a review of previous problems in the set involving a particular child.

Normally, a particular problem sequence was displayed two or more times, and T made two or more different responses to it before E moved to the next problem in the instructional sequence. It was not uncommon for E to refer to the standards established by the "jury of experts" which were used as a basis for rating T's performance and assessment of the problem. However, Es were instructed not to reveal the standard to T until he performed effectively in accordance with the standards. Also, E was instructed to continue instruction with a single problem sequence at least until T performed satisfactorily in accordance with the standards. E was allowed to repeat a particular problem sequence up to 10 times until necessary, but it was seldom necessary to repeat a problem more than five times. Normally, the total of 20 problem sequences were completed within three or four 50 minute sessions.

The post-test. Most Ts were post-tested during the week following the completion of instruction. The post-test procedure was the same as the pretest except that a new set of problems were employed, and all Ts were tested with life-size projections of the children. Each T was tested with the one set of 20 problems remaining out of the three sets used in the experiment.

In the few minutes remaining of the hour scheduled for post-testing, each T was asked to evaluate his experience in the simulated classroom by answering four questions:

1. How do you feel about your simulation experience?
2. Do you prefer the large as contrasted with the small projections of children? (This question was asked only of those who were exposed to both screen sizes.)
3. During instruction, we either projected motion pictures showing how the children would react to you or we simply described how they might react. In your case, we (state which feedback mode was used). Which method of providing feedback do you feel is most helpful?
4. Do you have any recommendations or suggestions for improving the instructional procedure?

The number of times T practiced each problem sequence during instruction also was used as a criterion measure. Theoretically, there could be a range of 180 repetitions (20-200 repetitions) because each T was allowed to practice each problem sequence up to 10 times. From previous research (Kersh, 1963), using different motion picture sequences but substantially the same instructional procedures, comparable Ts were observed to complete 20 problem sequences in an average of 43 practice trials (approximately two trials per problem). Furthermore, when still projections of children were employed instead of motion pictures, Ts required significantly fewer practice trials to complete instruction. From the findings of the previous study it was concluded that, when describing the problem to T as was required when still projections were used, E revealed important aspects of each problem directly to T. When motion pictures are used, by contrast, T is required to identify the same important aspects of each problem without help, and while being exposed to a great many more distracting cues from the motion picture projections.

Reliability of pretest and post-test. The performance test used in the present experiment as a pretest and post-test was the same that had been developed in connection with the previous research effort (Kersh, 1963). The motion picture sequences had been revised and new films made of a different class, but the problem sequences were in every other respect identical to the original ones. Only the children's names were different, and in many cases, the feedback sequences were changed or additional ones provided. These changes in no way altered the performance test procedures and standards because the feedback sequences were not used during testing.

The two research staff members who had originally been trained as experimental instructors (Es) were also employed in the present research. Consequently, interrater reliability estimates from the previous research effort also are applicable to the present. At that time, the two Es observed and rated 24 Ts simultaneously, and their independent ratings correlated +.68. For purposes of the present research effort, both Es simultaneously observed and rated a sufficient number of Ts during the pretesting period at the beginning of each quarter to be certain that they were still in close agreement. In effect, they were constantly striving together to improve the reliability of the rating procedure. Throughout the data gathering, the two Es consistently rated the same Ts independently within three total score points of each other.

Each term the two Es each instructed approximately one half of the available Ts, so that approximately the same number of Ts were instructed by each instructor in every treatment-combination group.

The reliability of the "index of learning rate" (number of practice trials per set of 20 problems) was estimated by comparing the mean number of practice trials per student for each of the two experimental instructors. This gross comparison indicated that the two Es differed in number of practice trials. One E tended to encourage his student teachers to practice more frequently than did the other. The mean number of practice trials was 40.35 for Ts instructed by the first instructor, and was 30.68 practice trials for students instructed by the second instructor. Subsequent investigation revealed that the first E tended to more frequently "rate" Ts during instruction on the basis of T's description of what he planned to do on the last practice trial. Consequently, the second E systematically eliminated what would normally be the last practice trial simply because it appeared redundant to do so after T had already verbalized how he would respond. A more complete analysis of instruction differences is reported in Chapter VI of this report.

Results

Mean pretest and post-test scores for each treatment combination group are shown in Table 1.1. Since Ts were assigned to treatment groups by random procedures, the observed differences in pretest means are presumed to reflect chance flucuations in performance. In fact, this assumption is supported by analysis of variance (not shown) which indicated that the differences are not statistically reliable. Consequently, differences in the post-test scores are presumed to reflect effects of the experimental treatments and treatment combinations.

Table 1.2 summarizes the analysis of variance of post-test scores. As is indicated, the observed differences in the main effects of orientation-pretest image size are statistically significant ($p < .01$), and the interaction of orientation-pretest size and feedback mode (OP x F) approaches significance at the .05 probability level.

The mean of the group which was pretested with small images was 43.20, and the mean for the group pretested with large images was 40.58. It may be presumed that employing small images in the pretest enhances instruction slightly. These findings, however, do not corroborate the findings of the first research project (Kersh, 1963) which indicated that small size projections during instruction also enhance learning. In the present experiment, differences between means of Ts instructed with the large and small size images were not statistically significant.

Observed differences in the pattern of means suggested that the OP x F interaction effect could well be an important one, so post-test scores were adjusted by analysis of covariance techniques, using pretest scores as the concomitant variable. This procedure was justified on the grounds that observed differences in pretest means do not reflect effects of the experimental treatment (orientation-pretest size). Rather, differences in pretest scores reflect individual differences in teaching competency as measured by the pretest.

Table 1.1

Mean Pretest and Post-test Scores for Experiment I Treatment Groups

Treatment Group	Pretest	Post-test
Large-large (vis)	29.90	37.90
Large-small (vis)	27.60	39.60
Small-large (vis)	28.50	43.60
Small-small (vis)	27.80	42.90
Large-large (ver)	29.00	43.00
Large-small (ver)	29.20	41.80
Small-large (ver)	27.70	43.50
Small-small (ver)	27.80	42.80

Table 1.2

Experiment I Analysis of Variance Summary: Post-test Scores

Source	df	MS	F
Orientation-Pretest Image Size (OP)	1	137.81	7.48**
Training Image Size (T)	1	1.01	—
Feedback Mode (F)	1	63.01	3.42
OP x T	1	4.52	—
OP x F	1	70.32	3.82
T x F	1	10.52	—
OP x T x F	1	10.50	—
Error	72	18.42	—
Total	79		

** $p < .01$

Adjusted post-test means for the various treatment combination groups are shown in Table 1.3. Actual changes in post-test means are minimal (see Table 1.1 for comparison), but the analysis of covariance summary in Table 1.4 does indicate that the interaction effects of orientation-pretest size and feedback mode is statistically significant ($p < .05$).

Figure 2 shows a profile of means for treatment combination groups involved. As is indicated, Ts who were oriented and pretested with the small images performed equally well in either the visual or verbal feedback mode. Ts who were oriented and pretested with large images, on the other hand, responded differently to instruction with visual and verbal feedback. Their scores on the post-test were slightly lower when they were instructed with visual feedback.

Responses to questions asked after post-test. The first question after the post-test asked for a general reaction from T, the second and third asked T's reaction to screen size and feedback mode, and the fourth asked for specific recommendation for improving the classroom simulation.

Responses to the first question are summarized in Table 1.5. Eighty Ts provided 295 responses, or an average of slightly less than four responses per T. Fourty-six Ts said they felt that the experience was generally beneficial, and 31 reported that they gained insight regarding the general classroom behavior of children. The other responses are recorded in order of frequency. It is noteworthy that, although a number of specific criticisms were leveled at the instructional procedure and mechanics of the laboratory facility, only one person felt that the experience was not generally beneficial.

Table 1.6 records the number who preferred large or small projections. Only the responses of those who had both large and small projections were recorded. As is indicated, two-thirds of those who had both large and small projections indicated that they preferred the large projections.

Table 1.3

Adjusted Treatment Group Means for Experiment I Post-test Scores

Treatment Group	Post-test
Large-large (vis)	38.21
Large-small (vis)	39.42
Small-large (vis)	43.61
Small-small (vis)	42.77
Large-large (ver)	43.12
Large-small (ver)	41.96
Small-large (ver)	43.34
Small-small (ver)	42.67

Table 1.4

Experiment I Analysis of Covariance Summary: Post-test Scores

Source	df	MS	F
Orientation-Pretest Image Size (OP)	1	114.93	6.37*
Training Image Size (T)	1	2.67	—
Feedback Mode (F)	1	62.64	3.47
OP x T	1	3.14	—
OP x F	1	76.15	4.22*
T x F	1	6.00	—
OP x T x F	1	8.04	—
Error	71	18.05	—
Total	78		

* $p < .05$

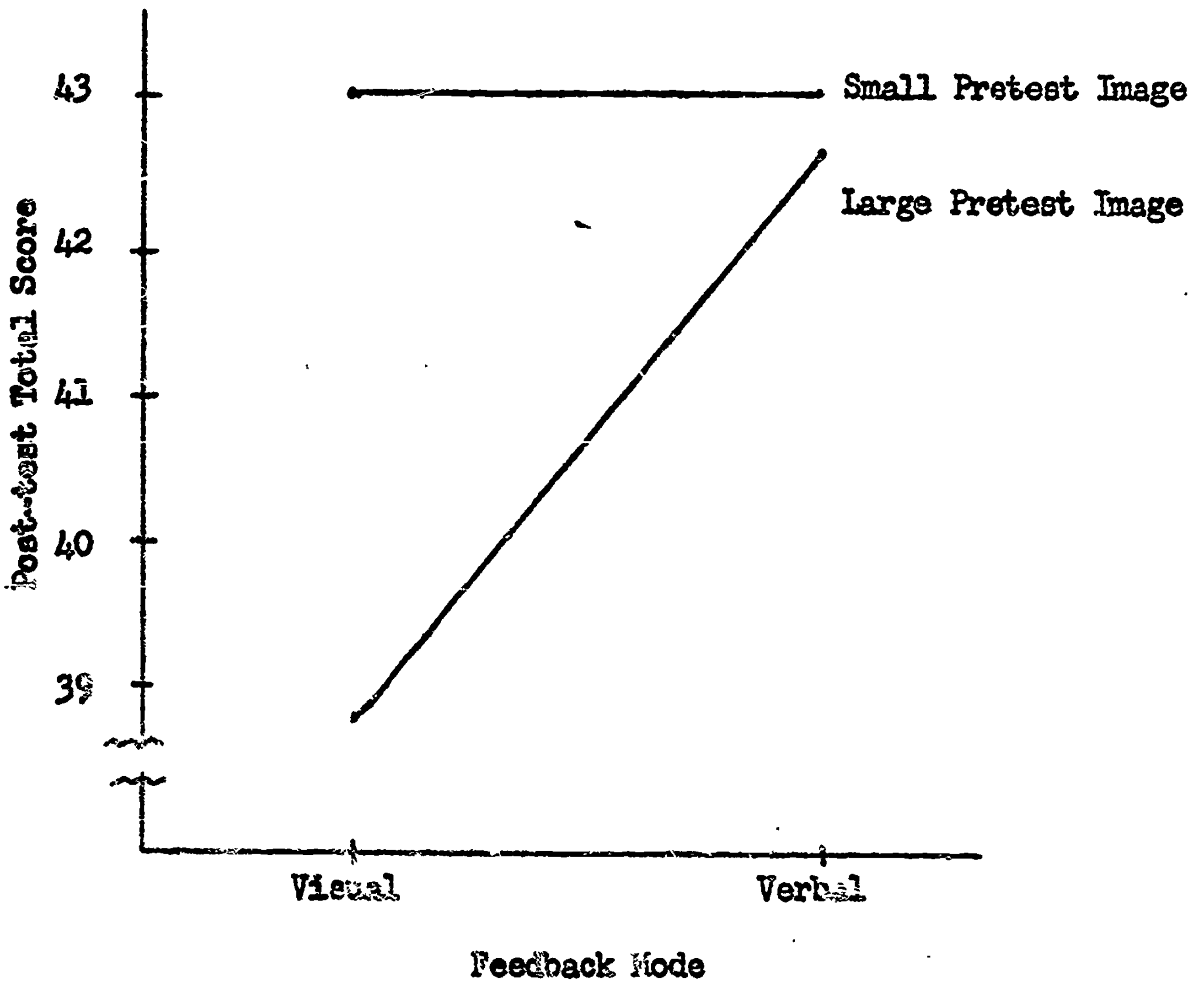

Figure 2. Profile of means showing interaction effect of orientation-pretest image size and feedback mode.

Table 1.5

Question 1. How do you feel about your simulation experience?

Responses	f
Generally beneficial	46
Gained insight regarding general classroom behavior of children.	31
Approve of "forced reaction" to problems.	29
Problems were realistic.	26
Children were realistic.	21
Felt nervous, threatened or frustrated.	20
Improved T's observational techniques.	19
Apprehensions diminished with experience in simulated classroom.	16
Problems were unrealistic.	12
Children were unrealistic.	11
Helped T's work in Education courses.	8
Gained general insight regarding classroom behavior of teacher.	8
Disapprove of "forced reaction" to problems.	7
Felt interaction with E was beneficial.	6
Had difficulty perceiving cues.	3
Helped build confidence	3
Had no difficulty perceiving cues	2
Approve of opportunity to practice handling situations different ways.	2
Generally not beneficial.	1
Miscellaneous comments by one T.	10
Comments which were not clearly recorded.	14
Total responses from 80 Ts	295

Table 1.6

Question 2. Do you prefer the large or the small projections of the children?

Response Category	f
Prefer large images	40
Prefer small images	11
No preference	9
Not recorded (Large-large treatment groups)	20
Total responses	80

Table 1.7 summarizes the responses to questions concerning the visual and verbal feedback. A total of 51 out of 80 Ts indicated that they preferred the visual and 14 preferred the verbal. The remainder had no opinion or expressed no comment. Notice, however, that there is a tendency for Ts to prefer the particular mode in which they were instructed. The visualized feedback was occasionally interspersed with verbal feedback because the feedback sequences on film were not always appropriate for particular responses to the problem. Nevertheless, there was a distinct majority preference for visual displays.

Recommendations generally for improving classroom simulation are summarized in Table 1.8. A total of 141 responses were recorded for the question, or an average of slightly less than two responses per person. Although interesting, the distibution of recommendations is relatively large with no recommendation being particularly outstanding.

Table 1.7

Question 3. Which method of providing feedback do you feel is most helpful?

Response Category	f
A. Those having visual feedback	
Prefer visual	33
Prefer verbal	1
No opinion	1
No comment recorded	5
B. Those having verbal feedback	
Prefer visual	18
Prefer verbal	13
No opinion	5
No comment recorded	4
Total	80

Table 1.8

Question 4. Do you have any recommendations or suggestions for improving the instructional procedure?

Recommendations	f
A. Prior to Training	
Provide more orientation experience	15
Explain more about technique to T	5
Provide more background on class	3
Eliminate pretest	2
Provide more information on sequence of problems	2
Orient and pretest on separate days	2
Make records more accessible	1
Explain rating system	1
Test individually	1
Provide prior classroom experience	1
Simulate interview with Mr. Land	1
Provide more information on T-supervising teacher relationship	1
Check-out orientation slides	1
B. Training Situation	
Discuss pretest sequence (problems)	3
Give T knowledge of success during pretest	2
One E only	1
Provide more discussion about problems	1
More review	1
Advise Ts to review folders	1
Recommend more visual feedback	1
C. Place in Educational Program	
Use with "Block" courses	13
Use with sophomore courses	9
Use in groups	2
Use with junior courses	1
Use to replace block observation	1
D. Mechanical	
Improve seating chart pictures	3
Improve sound	3
Vary screen size	1
Make screen larger	1
Use only large screen	1
T should be able to stop film	1
Training facility should be larger	1

Table 1.8 (Cont'd.)

Recommendations	f
E. Miscellaneous	
I would recommend it to others	27
Gave no recommendation for improvement	14
Ambiguous or incomplete statements	5
Provide films for different age levels	3
Provide films for other school activities	3
Combat bad image	1
Don't call tests "tests"	1
Liked choosing alternate responses	1
Provide contact with the real class later	1
Liked indirect questioning of E	1
Combat leakage of problems and solutions	1

Discussion

In simple terms, answers to the questions posed before the experiment are as follows, (based on present findings):

1. Yes, groups which are oriented and pretested in the large projection mode differ in the post-test performance from those which are oriented and pretested in the small project modes. Those which are oriented and pretested in the small projection mode tend to score slightly higher on the post-test.
2. No, groups which receive visual feedback during instruction do not differ markedly in their post-test performance from those who receive verbalized feedback.
3. No, groups which are instructed in the large projection mode do not differ markedly in their post-test performance from those which are instructed in the small projection mode.
4. Yes, there are particular interaction effects which occur when the instructional treatments are combined. Groups pretested in the large projection mode perform slightly better on the post-test when instructed with verbalized feedback (as contrasted with visual feedback). Groups pretested with small projections perform equally well on the post-test after instruction with either feedback mode.

The findings indicate that less realistic (small) projections result in higher post-test performance scores than life size (realistic) displays. However, observed differences between treatment groups are very small even though statistically significant.

If existing evidence were used as a basis for developing a simulation facility which would maximize instruction (recognizing that additional gains in performance would be very small at best), findings from present and previous studies would support the use of small projections throughout orientation and instruction. Thus, the instructional media could be made adaptable for use in a much less elaborate and less expensive laboratory facility. Conceivably, an instructional facility could be developed for use with eight mm. motion picture projectors, perhaps of the newer cartridge-loading, self-threading type, so that the student teacher himself could manipulate the controls and replace film cartridges in accordance with written instructions. Conceivably, instruction could be provided to several student teachers simultaneously with equal effectiveness, and at considerable reduction in operating costs. Instruction could also be group-paced.

Experiment II, reported in the following section, tests the importance of T overtly enacting responses to the motion picture sequences as contrasted with simply describing what he would do. If overtly enacting responses to the problem sequences is also of little consequence, further development of instructional media along the lines indicated above would receive further justification.

In the introductory sections of this present report, it was argued that the use of life size projections during orientation and pretesting followed by less realistic small projections during the instructional phase could result in increased learning. It was originally argued that the sequential effect of large followed by small projections would first of all motivate the student, then enable him to practice under conditions which would be more conducive to learning. Present findings are not supportive to this explanation of the previous findings.

V Experiment II: Response Mode

The second experiment was designed to answer the following question concerning T's mode of response during instruction.

> Do Ts who enact responses to the problem on film differ in their post-test performance from those who simply describe how they would respond?

Responding overtly was considered to be a more realistic mode than verbally responding to the problem because the enacted response more nearly approximates the response T makes during actual instruction. The experiment compared one treatment against the other, using post-performance in the simulation facility as the criterion.

Experimental Design

Data gathering was continued as in Experiment I with Ts selected from the junior block course at Oregon College of Education during winter and spring quarters, 1965. Data gathering continued until a minimum of 15 Ts had completed instruction in each of two treatment categories identified as follows:

1. Enacted response. Individualized instruction was provided with each T enacting his response to each problem sequence.
2. Verbalized response. Individualized instruction was provided, but with each T describing rather than enacting his responses to the problems.

Selection of subjects. Subjects were screened on the basis of age and previous teaching experience as in Experiment I, but the full range of pretest scores was included. Instead of orienting and pretesting each T individually, group procedures were employed. The following steps were taken each term in assigning Ts to experimental groups:

1. All Ts in the junior block course were ranked by College Entrance Exam Board scores and sex as in Experiment I.
2. Each T was assigned to one of three groups by table of random numbers, and each group was pretested by a different set of problem sequences (Program I, II, III).
3. Ts were oriented and pretested in groups, using specially prepared response sheets to record their responses. The timing of their response to each problem was indicated by a record of elapsed time recorded by each T. A timing device was employed which projected numbers on the screen at two second intervals while each problem sequence was being projected.
4. The written responses were scored using revised scoring standards (referred to in Chapter II). Where timing of response served as a criterion for scoring, the record of elapsed time recorded by Ts was used.
5. Ts were assigned to one of the two experimental treatment groups by a table of random numbers.
6. Ts were assigned a meeting time for individualized instruction in the simulation laboratory, and the two Es alternated instruction throughout the week in three hour blocks. Ts were not given a choice of instructor.

Changes in instructional procedure. The revised instructions and standards referred to in Chapter II were employed in Experiment II. In addition, a slightly different instructional procedure was employed. In the second experiment, instruction was continued until T made at least two responses to each problem, one representing the response considered most effective by the rating standards, and the other response which could be rated as least effective. E encouraged T to give alternative responses without giving information directly concerning the response standards. For example, E was instructed to ask such questions as, "Can you think of another way to handle the situation?" or, "Now that you have seen the problem and thought about it, can you think of a different way to handle it?"

Finally, after T had made at least two responses, one at either extreme by the rating standards, E was instructed to "reinforce" the response recommended by the rating standards. This was done by asking T to select the response he would consider most effective. If T chose the "wrong one," E was instructed to say, "The experts who developed our standards would suggest (description of response)." E also was instructed to avoid explaining the rating standards to T unless he asked for an explanation. If T asked for an explanation, E gave it in terms of the specific problem under consideration, e.g., "If you were to interrupt the class and speak to Brian about tapping his pencil, it would disturb the others and thus might not be so effective." E was instructed to avoid simply restating the standards as they were printed on the instruction sheet.

Those instructed by the verbalized response treatment were instructed exactly as those who enacted their response except that they were asked to describe rather than to enact their responses to the motion picture sequences. Specifically, each T in the verbalized response group was instructed to look at the motion picture sequence until he thought something should be done about the situation, then to signal E to stop the forward motion of the projector (so that a single frame was projected) and to describe to E what he should do. After listening to T's response, E projected the appropriate feedback sequence.

Post-test and data analysis. Each T was post-tested individually as in Experiment I, except that the revised rating standards were employed. The revised rating procedure employs more objective techniques for assigning separate numerical values to T's response ("R" score) and assessment

("A" score) of each problem sequence. Vlcek (1965), using the same rating procedure, had previously reported no statistically significant difference between his experimental and control groups using A scores. Only R scores effectively measured learning from classroom simulation training. Twelker, in a recent analysis of classroom simulation data from another study, found that R scores are a more sensitive measure of treatment effects.[1] Consequently, the decision was made to analyze R scores, A scores, and total scores separately in the analysis of Experiment II data.

Reliability of the revised post-test. The reliability of the revised post-test has been estimated by Vlcek (1965) and Twelker.[1] Vlcek reported inter-rater reliability coefficients of .91 (n = 12 observations of four Ts) and .87 (n = 18 observations). However, at the end of the college quarter, findings from his post-test data indicated that the reliability had dropped to .63 (n = 195 observations of 10 Ts). Vlcek's findings underline the importance of periodic retraining of Es as was the procedure in the present research.

Twelker's estimates were all based on data collected immediately after his Es were trained. Using R scores only from 20 Ts, he pooled data from four raters by an analysis of variance technique (Winer, 1962, pp. 124-32) and obtained a coefficient of .94.

---

[1] Personal communication with Paul Twelker, Assistant Research Professor, Teaching Research Division, 1965.

Results

Although the experimental subjects were not assigned to treatment groups by pretest levels, a sufficient number of subjects were instructed to permit equal samples to be drawn from those scoring above and below the median pretest score for all experimental subjects. It may be remembered that the full range of pretest scores were included in the sample for Experiment II, in contrast with the procedure employed in Experiment I wherein those scoring above a pretest score of 34 were eliminated.

The median pretest "R score" for the entire sample was 35, and the range was from 26 to 49. Eight Ts were selected who scored above and below the median for each of the treatment combination groups. The four treatment combination groups were identified as follows:

1. High verbal. Ts scoring above the median on pretest, and instructed by verbal response procedure.
2. Low verbal. Ts scoring below the median on pretest, and instructed by the verbal response procedure.
3. High enacted. T scoring above median on pretest, and instructed by enacted response procedure.
4. Low enacted. T scoring below median on pretest, and instructed by enacted response procedure.

Tables 2 - 2.5 present the results of the statistical analysis. Clearly, the observed differences among means are not statistically significant.

Table 2

Experiment II Mean Post-test P Scores

(n = 32)

Pretest Level	Response Mode Verbal	Enacted
Above Median	42.50	41.75
Below Median	37.75	43.00

Table 2.1

Experiment II Analysis of Variance Summary: R Scores

Source	df	MS	F
Pretest levels (L)	1	24.5	—
Response Mode (R)	1	40.5	1.45
L x R	1	72.0	2.58
Error	28	27.9	—
Total	31		

Table 2.2

Experiment II Mean Post-test A Scores

(n = 32)

Pretest Level	Response Mode Verbal	Enacted
Above Median	27.38	27.38
Below Median	25.00	25.25

Table 2.3

Experiment II Analysis of Variance Summary: A Scores

Source	df	MS	F
Pretest levels (L)	1	40.50	3.28
Response Mode (R)	1	0.13	—
L x R	1	0.13	—
Error	28	12.33	—
Total	31		

Table 2.4

Experiment II Mean Post-test Total Scores

Pretest Level	Response Mode Verbal	Enacted
Above Median	69.88	69.13
Below Median	62.75	68.25

Table 2.5

Experiment II Analysis of Variance Summary: Total Scores

Source	df	MS	F
Pretest levels (L)	1	128.00	2.76
Response mode (R)	1	45.13	—
L x R	1	78.13	1.68
Error	28	46.46	—
Total	31		

Discussion

Findings from Experiment II reported above, indicate that there is no statistically significant difference in post-test performance of Ts who enact responses to problems on film and those who simply describe how they would respond. This finding adds further support to the suggestion made in the discussion of Experiment I findings, that classroom simulation as an instructional medium may be adapted to individualized or group-paced instruction where Ts view smaller than life size images and respond by describing rather than enacting what they would do.

Inspection of pretest and post-test differences (gain scores) indicated that a pattern of differences might exist for below-median pretest groups that would be different for above-median pretest groups. Furthermore, from findings in Experiment I that one E allowed his Ts to practice fewer times than did the other E, a more extensive analysis of the data in terms of the gain scores and instructors was conducted. This analysis is reported in the next section.

VI Analysis of Practice Effects and Instructor Differences

The reader is reminded that the instructional procedure in Experiment I and Experiment II, although slightly different, had one characteristic in common. Ts were allowed to practice responding to each problem until their performance could be assigned the highest rating by the standards employed. In effect, each T was instructed in the simulation laboratory until he attained a prescribed behavioral standard. It could be expected that those scoring low on the pretest "gained" more from instruction than high-scoring individuals. Although the Experiment II analysis compared post-test performance of Ts scoring high and low on the pretest, the analysis of Experiment I data did not. Also, the analysis of Experiment II data did not consider pre-to- post-test increments (gains). Further analysis of Experiment I data in terms of pretest levels is possible, even though Ts were not assigned to experimental groups with this in mind.

Another finding which warrants further investigation is that reported in an earlier section (see Experiment I) that Ts instructed by E_1 practiced more than Ts instructed by E_2.

Data from Experiment I and Experiment II were analyzed further in an effort to answer the following questions:

1. Did those who scored low on the pretest take more practice trials to achieve the instructional criterion than did those who scored relatively high on the pretest?

2. Did those who scored low on the pretest in fact "gain" more than those who scored high initially (as measured by differences between pretest and post-test scores)?

3. Did the fact that E_1 allowed his students to practice more than did E_2 result in consistently greater gains for those trained by E_1?

Analysis of Practice Trial Data

The number of times each T practiced responding to each problem during instruction was recorded for Experiment I but not for Experiment II. In the analysis of these data reported earlier (Chapter IV, Experiment I) it was reported that E_2 consistently instructed his Ts with fewer practice trials than did E_1. Specifically, in Experiment I, Ts taught by E_1 practiced an average of 40.35 times during the instructional phase, whereas those taught by E_2 practiced 30.68 times. However, no effort was made to ascertain whether or not Ts scoring low on the pretest required a greater number of practice trials to reach the criterion than did those who scored high initially. In order to analyze the practice trial data by pretest levels, and at the same time to take into account instructor differences, it was necessary to collapse treatment groups and select a small sample from the original group of 80 Ts who participated in Experiment I.

Equal numbers of Ts scoring above and below the median pretest score were selected from those taught by each E and who were oriented and pretested by each of the two image sizes. The orientation-pretest image size treatment groups were included in the analysis because the effects of this experimental treatment were shown to be statistically significant in the previous analysis. The resulting sample numbered 48, with 5 and 7 Ts distributed proportionately among the 8 treatment-combination groups.

Table 3 summarizes the analysis of variance for the practice trial data. Only differences between groups taught by the two instructors are statistically significant. Table 3.1 includes the mean number of practice trials for each treatment group.

Table 3

Analysis of Variance Summary: Experiment I Practice Trial Data

Source	df	MS	F
Pretest Level (L)	1	24.08	0.64
Orientation-pretest image size (OP)	1	2.08	0.06
Experimental instructor (E)	1	1091.26	29.20**
L x OP	1	10.08	0.27
L x E	1	94.29	2.52
OP x E	1	7.20	0.19
L x OP x E	1	138.29	3.70
Error	40	37.37	--
Total	47		

** $p < .01$

Table 3.1

Mean Number of Practice Trials for High and Low Pretest Groups: Experiment I Data

($n = 48$)

Pretest Group	Instructor E_1 ($n = 10$ per group)	Instructor E_2 ($n = 14$ per group)	$E_1 + E_2$
Below median	39.40	32.57	35.42
Above median	41.30	28.79	34.00
Combined	40.35	30.68	

Analysis of Gain Scores

Apparently, Ts who scored low on the pretest did not require more practice than the high-scoring Ts to attain the instructional criterion. Did low-scoring Ts "gain" more as measured by differences between pretest and post-test scores? Gain scores were obtained by subtracting pretest scores from post-test scores for all Ts selected for the analysis recorded above (n = 48). Table 3.2 summarizes the analysis of variance for the gain score data. As is indicated, the difference between means of the groups at each pretest level are significant, and also there is a significant interaction effect involving the two instructors and the two pretest groups.

Table 3.3 shows the means for each treatment combination group taught by the two Es at each pretest level. The pattern of gain scores is more clearly evident in Figure 3. Figure 3 indicates that those scoring below the median on the pretest did, in fact, "gain" more from instruction than did those who scored above the median; however, the low pretest group taught by E_1 gained more than did the comparable group taught by E_2, and the opposite was true for the high pretest groups. Apparently, E_1 was more effective with the low pretest group and E_2 was more effective with the high pretest group.

In order to reveal this interaction effect more clearly, post-test total scores of the same group (n = 48) were reanalyzed. Table 3.4 summarizes the analysis of variance for post-test scores, and indicates the same interaction effect to be statistically significant. Table 3.5 shows the mean post-test total scores, and Figure 4 illustrates the interaction effect graphically. The crossed pattern clearly reveals that E_1 was a more effective instructor with Ts in the low pretest group, and E_2 was more effective with Ts in the high pretest group.

Table 3.2

Analysis of Variance Summary: Experiment I Pre- Post-test Gain Scores

(n = 48)

Source	df	MS	F
Pretest Level (L)	1	320.33	11.99**
Orientation-pretest image size (OP)	1	44.08	1.65
Experimental Instructor (E)	1	9.45	0.35
L x OP	1	16.33	0.61
L x E	1	125.95	4.71*
OP x E	1	2.29	0.09
L x OP x E	1	0.24	0.01
Error	40	26.71	
Total	47		

* $p < .05$
** $p < .01$

Table 3.3

Mean Pre- Post-test Gain Scores for High and Low Pretest Groups: Experiment I Data

(n = 48)

	Instructor	
Pretest Group	E1	E2
Below median	17.40	13.21
Above median	8.40	10.79

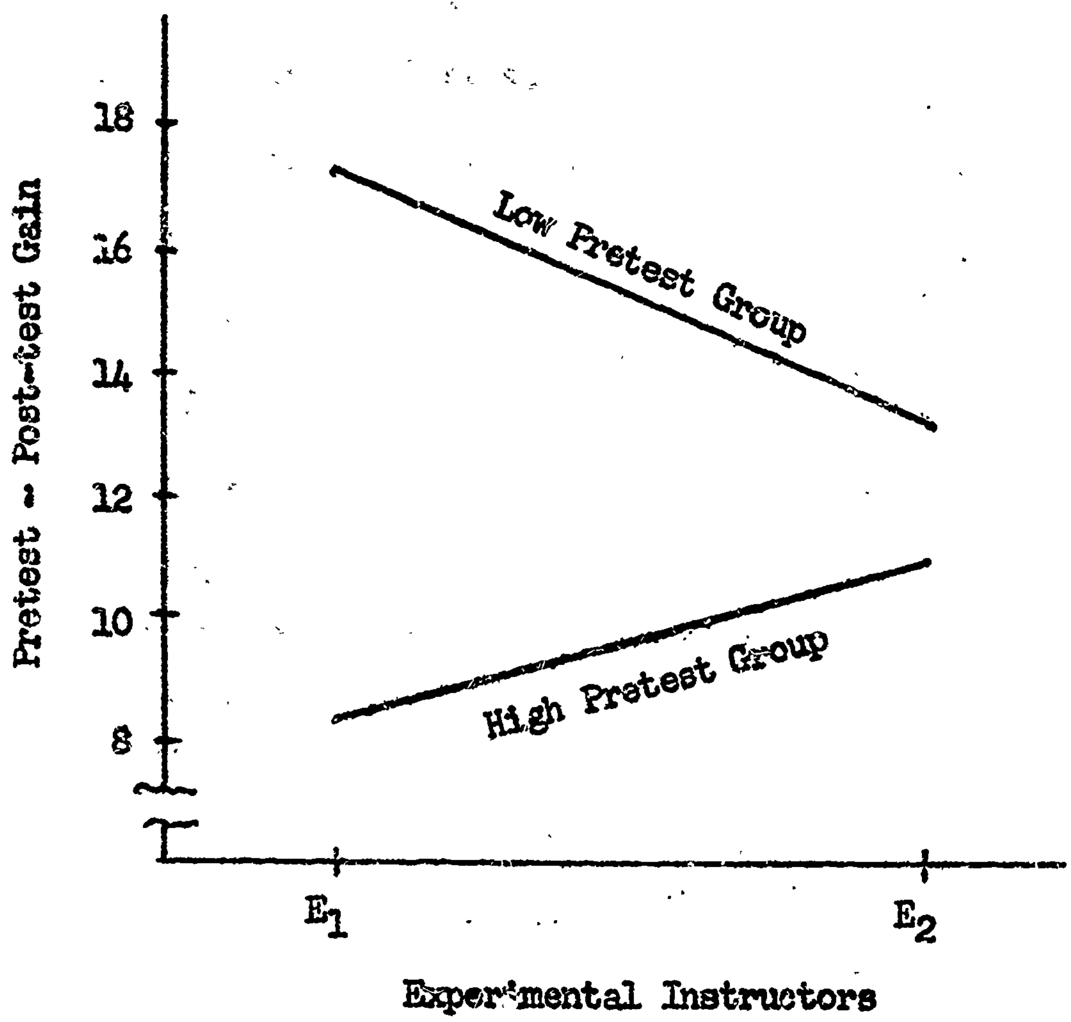

Figure 3. Instructor differences with high and low pretest groups using "gain" scores as criterion.

Table 3.4

Analysis of Variance Summary: Experiment I Post-test Total Scores

(n = 48)

Source	df	MS	F
Pretest level (L)	1	4.69	0.23
Orientation-Pretest Image Size (OL)	1	54.19	2.68
Experimental Instructor (E)	1	2.83	0.14
L x OP	1	1.69	0.08
L x E	1	89.15	4.41*
OP x E	1	0.90	0.04
L x OP x E	1	1.55	0.08
Error	40	20.23	
Total	47		

* $p < .05$

Table 3.5

Mean Post-test Total Scores for High and Low Pretest Groups: Experiment I Data

(n = 48)

Pretest Group	Instructor E1	Instructor E2
Below median	42.90	39.64
Above median	40.30	42.57

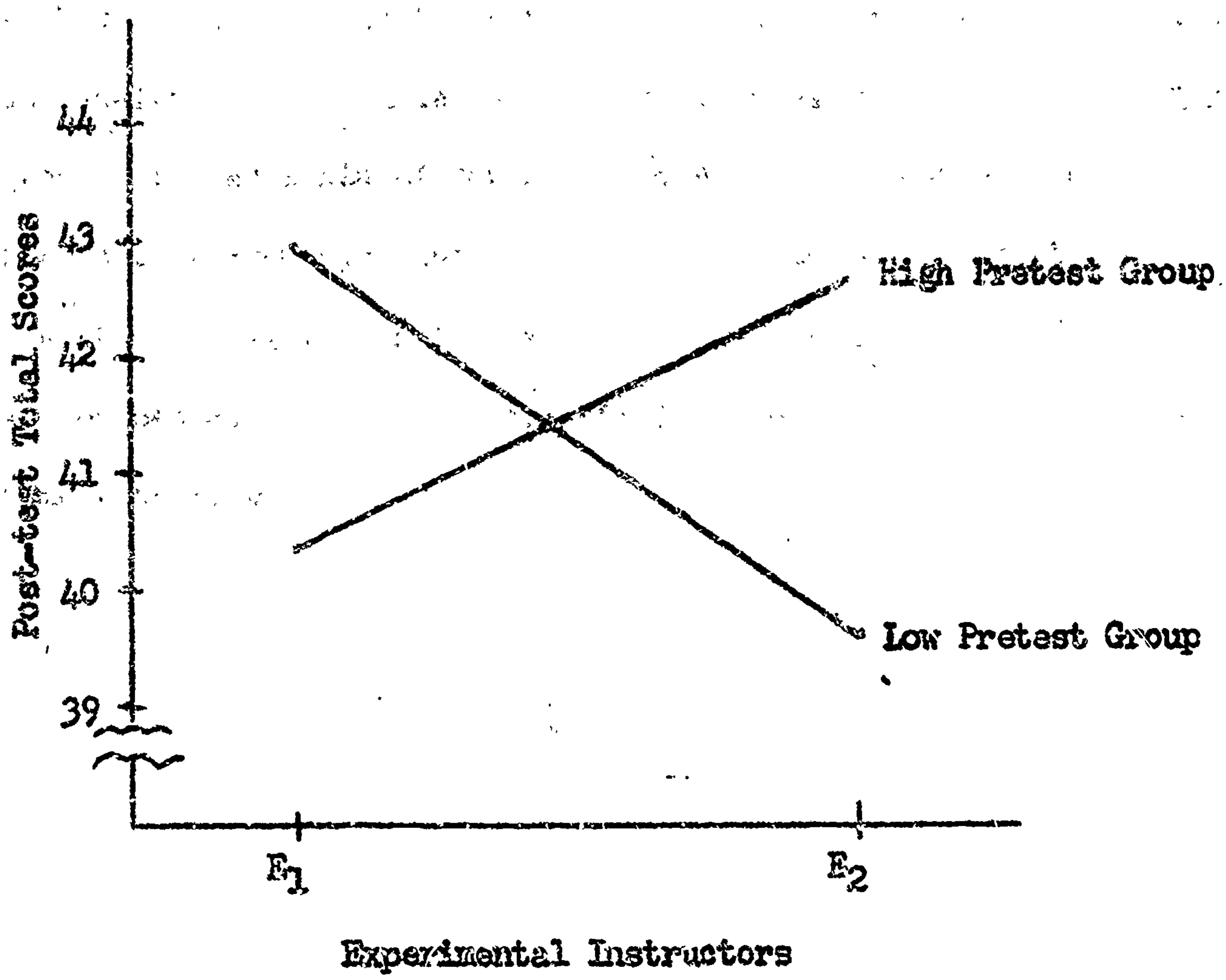

Figure 4. Instructor differences with high and low pretest groups using post-test performance ratings as criterion.

Although a modified rating procedure was employed in Experiment II, and Ts were not screened on the basis of pretest performance, the gain scores for Experiment II were subjected to a similar statistical analysis. Treatment groups were combined, and equal numbers above and below the median pretest score were selected who were taught by each E. From the total sample of 37 Ts involved in the data gathering for Experiment II, a sample of 36 was selected, with 9 in each of the 4 treatment combination groups. Table 3.6 summarizes the analysis of variance, and Table 3.7 presents the means for each treatment group. The analysis of variance indicates that the main effects of differences between high and low pretest groups and groups taught by E_1 and E_2 are statistically significant. The interaction of Es and pretest levels approaches significance at the .05 level closely enough to be acceptable by some researchers.

Table 3.6

Analysis of Variance Summary: Experiment II Gain Scores

(n = 36)

Source	df	MS	F
Pretest level (L)	1	702.25	18.52**
Experimental Instructor (E)	1	182.25	4.81*
L x E	1	156.25	4.12
Error	32	37.92	—
Total	35		

* $p < .05$ ($F = 4.15$ for 1, 32 df)
** $p < .01$

Table 3.7

Mean Pre- Post-test Gain Scores for High and Low Pretest Group
Experiment II Data

(n = 36)

Pretest Group	Instructor E1	Instructor E2	E1 + E2
Below median	11.22	10.89	11.05
Above median	6.56	-2.11	2.23
Combined	8.89	4.39	

Discussion

Based on the analysis of practice trial data and gain scores, answers to the questions raised at the beginning of this section may be generalized as follows:

1. Student teachers whose performance in the simulated classroom is relatively low initially do not require more practice to learn to perform to a prescribed standard than student teachers whose initial performance is relatively high.

2. Given equivalent individualized instruction in the simulated classroom, Ts whose initial performance is rated relatively low gain more from instruction than those whose initial performance is rated higher, so that, after instruction, all Ts perform at approximately the same level.

3. Using identical instructional materials, different instructors trained to follow a given instructional procedure are not equally effective with students at differing initial performance levels. In other words, given two instructors trained to use a particular instructional procedure, one may prove to be more effective with Ts scoring relatively low on a pretest of performance, and the other may prove to be more effective with Ts who score relatively high on a pretest. Instructor differences may not be attributed entirely to individual variations in amount of practice provided each T during instruction.

Considered in the context of previous findings concerning such instructional variables as image size of projections, response mode employed by the student teacher, and the use of motion pictures as contrasted with still projections--all of which have to do with the fidelity of the simulated setting, i.e., how accurately the simulated setting approximates the "real" setting--the findings from the last analysis indicate that there are far more powerful instructional variables to be considered. The fact that Ts who perform relatively low initially are more responsive to instruction in the simulation laboratory than those who are initially rated higher (that is, low-scoring Ts "catch up" with and ultimately perform as well as those who are higher in their initial performance) suggests that the particular type of individualized instruction provided in the present research may be uniquely fitted to the needs of low-performing student teachers. Those whose performance in the simulated classroom is rated to be in the middle or higher range of the pretest may respond as well from other forms of simulation training, e.g., procedures which are group-paced or which do not require direct supervision. Such training modes would be relatively less expensive and could be adapted for use on a wide scale.

The classroom simulation materials described in this present report are being modified and adapted for use with groups or for relatively unsupervised use by student teachers individually. When such materials and procedures are available, further research may be conducted to determine whether or not persons whose initial performance is rated low do, in fact, respond more readily to the type of instruction employed in the present research, and comparisons may be made with individuals whose initial performance is rated higher.

Another question worthy of further investigation is that which pertains to instructor differences. The two instructors in the present research, although trained to use a given instructional procedure did have a differential effect on their students. The only recorded difference in their procedure was that which pertained to a number of practice trials. One E tended to practice his student teachers more frequently than did the other. However, this difference in practice was not reflected in differences between the two groups of students taught by the two instructors. Instead, the two instructors were more or less effective with students in differing pretest groups. It may be that individualized methods of interacting with students during the discussions of each problem sequence account for instructor differences.

VII Follow-up Studies[1]

Does simulation training transfer to the actual classroom? This question is of primary concern, but as yet no effective means for measuring transfer effects has been devised. Previous studies by the present investigator have relied on a performance test in the simulation laboratory using a set of simulated problems different from those employed during instruction. Although this post-test of performance in the simulation laboratory is a test of transfer, it certainly is not so convincing as would be an assessment of T's performance in an actual classroom situation. If simulation training cannot measurably improve T's performance in the classroom, its value in teacher education programs could be challenged.

At Oregon College of Education, students enrolled in junior block have a great variety of experiences with the children in classroom situations, and they participate as instructors for short periods of time, occasionally in connection with professional courses which follow. However, the first occasion when all students may participate as instructors with full responsibility for an entire class follows simulation training by at least one year. It would be expected that any improvement brought about from five hours of simulation training would be most difficult to measure after so long a period of time.

Another difficulty in measuring the transfer effects of simulation training is that there is no observational technique known to this investigator that would measure the specific skills and techniques learned in "Mr. Land's Sixth Grade." It would be necessary to devise a completely new observational technique or to modify an existing observational procedure.

---

[1] The Follow-up Studies were carried out under the direction of Paul Twelker, Assistant Research Professor, Teaching Research Division.

Finally, because nearly every student teacher enrolled in the junior block course at Oregon College of Education was provided some form of simulation instruction, it was difficult to identify a "control" group for comparative purposes. Students enrolled in other courses, such as Educational Psychology, which could be substituted for junior block did not have the same pattern of practicum experiences.

All of these factors made the study of transfer effects of simulation instruction extremely difficult, if not impossible. Nevertheless, as a pilot effort in connection with Experiments I and II reported previously, an attempt was made to follow up on Ts who completed simulation instruction and to observe their performance in actual classroom situations during their teaching experience. The primary objective of these so called "follow-up studies" was to explore different techniques for observing and recording the classroom behavior of student teachers as they encountered problems similar to those included in the simulation materials. The follow-up studies could not be expected to provide sound evidence concerning the transfer effects of simulation instruction.

As reported in a previous section (Related Research), Vlcek (1965) used what was described as a greatly modified Flanders' technique to observe Ts at Michigan State University who enrolled in student teaching the semester immediately following their experience in the simulated classroom. Actually, Vlcek's observational technique bears little resemblance to Flanders' Interaction Analysis technique. Vlcek's observers used a specially prepared observation form to record events in the classroom. They observed problem situations in the classroom which corresponded to those included in the simulation experience, classified them by type, and then recorded how T responded. Vlcek's observers also evaluated how T's pupils reacted.

Vlcek reported that student teachers who completed simulated instruction acted in accordance with a greater number of the teaching principles than did the control group, but not necessarily in response to the types of problems included in the instructional materials. From these data Vlcek concluded that student teachers do learn principles of instruction which are transferred to the classroom, but that they do not necessarily employ them effectively (as judged by the pupils' response to T's efforts).

A modified Flanders' Interaction Analysis technique was also employed in follow-up studies for the present project, but the observational procedure followed closely that which Flanders has developed (Flanders, 1964). Modifications were based on an analysis of the behaviors learned in Mr. Land's Sixth Grade, the simulated classroom. From the analysis it was concluded that classroom interactions between teacher and pupil which are judged to be "private" (i.e., interactions with only one pupil) should be recorded separately from those which are judged to be "public," (i.e., wherein T speaks to more than one pupil at a time, or speaks to a single pupil loudly enough to be heard by others in the vicinity). In addition, it was considered important to record whether a teacher-pupil interaction is preceded by T "moving in" to communicate "privately" with a pupil, or communicated from a distance (without moving in). Although it was recognized that many circumstances would allow for exceptions to the general rules or standards taught in simulation instruction, it was nevertheless reasoned that Ts with experience in Mr. Land's Sixth Grade might tend to more frequently "move in" and communicate at close range (i.e., "privately") following a pupil-initiated "private" interaction.

To facilitate recording and analysis, the first two categories of teacher behavior employed in Flanders' category system ("Accepts Feeling," and "Praises or Encourages") were combined and recorded in the second category.

Whenever T was observed to "move in" to close range when communicating with a pupil or to a group of pupils, the first row or column was employed in the interaction matrix. Also, instead of adding categories to those developed by Flanders, intercommunications which were judged to be "private" as contrasted with "public" were indicated by a subscript "i" (private) or "p" (public) to the numerals which designate the observed behavior.

For example, consider a hypothetical problem situation in T's classroom. During a study period, two boys are observed to begin talking together and it appears that one is explaining something to the other (Category 9, "Student Initiated Behavior"). However, they talk loudly enough to be disturbing to others. If T "moves in" (Category 1), and speaks softly to the boys, directing them to lower their voices (Category 6, "Giving Directions"), the interaction would be recorded as follows: 9-1-6i. If, on the other hand, T is observed to direct the boys from across the room loudly enough to be distracting, the interaction sequence would be recorded as follows: 9-6p.

It was anticipated that patterns could be identified from records of classroom interactions and used as evidence of transfer. A "9-1-6i" pattern would be an interaction pattern expected of Ts who completed simulation training, and a "9-6p" pattern would be used as a negative instance.

Other classroom observational techniques were considered, including the relatively new observation system developed by Schalock, Beaird, and Simmons (1965). However, the observational categories used in each system were not directly related to the principles supposedly taught by simulation training, and the modified Flanders' technique most nearly met the requirements of the present study.

The principle investigator together with a research associate trained together using motion pi ure sequences of classroom scenes as stimulus materials, and subsequently by observing together in classroom situations.

The effort was to achieve an interrater reliability at the level recommended by Flanders. Flanders recommends a Scott coefficient of .85 or higher (Flanders, 1964, p. 15). Initially, the two observers had great difficulty in achieving and maintaining the recommended level of agreement. In an effort to increase the interrater reliability estimates, special battery-powered metronomes were designed and utilized which signal E every three seconds by tactile stimulus to the hand. This procedure added precision but did not significantly increase interrater reliability estimates. Only by pooling larger numbers of observations, was it possible to approximate the levels recommended by Flanders.

The most discouraging aspect of the observations, however, was that detailed records of T's performance in the classroom for a total time of approximately 60-90 minutes (20-30 minutes on three different occasions) simply did not provide sufficient basis for evaluating transfer. To obtain more extensive records was not within the scope of the present pilot effort. After an intensive and systematic observational study of approximately 20 Ts, the decision was made to abandon direct observation as a technique for data gathering.

Finally, a questionnaire was developed to be completed by T's supervising teacher after observing T's classroom behavior in a wide variety of circumstances over a period of weeks. The questionnaire presented questions that pertained directly to standards taught in simulation training. Admittedly, many pitfalls accompany the use of questionnaires. However, it was felt that if the questionnaire was administered carefully, meaningful data could be gathered which would provide more pertinent evidence than that which could be obtained through direct observation over limited periods of time.

Selection of Subjects

As was pointed out above, a primary difficulty in studying the transfer effects of simulation training (in the present study) was the lack of a control group. Only a few Ts completed the junior block course without having also been trained in the simulated classroom. Fortunately, a single group of approximately 30 student teachers did complete the course prior to Experiment I, without having participated in simulation training. Follow-up data were obtained approximately one year later on 11 student teachers who had undergone simulation, and 10 from the previous class which had not undergone training.

Supervising Teacher Questionnaire

The Supervising Teacher Questionnaire (see appendix) consisted of 17 questions. Fifteen questions involved behaviors directly related to the standards employed in classroom simulation training. The remaining two questions asked for a general assessment of the student teacher pertaining to (1) his ability to quickly assume responsibility for the class, and (2) the types of problems with which the student teacher had difficulty. Each supervising teacher was instructed by E in the use of the questionnaire.

One question which illustrates the type included in the questionnaire is the following:

> When direct action was required to control a disruptive group or individual, did the student teacher "move in" and communicate at close range as compared with communicating at a distance (e.g., yell over noise or disrupt a quiet study period)?

In response to such a question, the supervising teacher was asked to indicate (1) how frequently T responded in the manner indicated, and (2) how frequently a problem arose which required T to use such a technique.

The supervising teacher used a five-point scale ("always" to "never") to respond. For example, a supervisor might estimate that his student teacher "moved in" every time the opportunity occurred ("always"), yet might indicate that the problem occurred rather rarely ("seldom").

In responding to the question regarding how quickly T was capable of assuming full responsibility for the class, the supervising teacher responded by checking a given category indicating the number of weeks after the beginning of student teaching (within 1 week, 1-3 weeks, 4-6 weeks, 7-9 weeks, after 9 weeks).

Results

Each point on the five-point scale was assigned a value, and the data were analyzed by parametric techniques. Means and standard deviations for the experimental and control groups for each of the fifteen questions involving frequency of certain behaviors, and the sixteenth question involving speed of adjustment, are shown in Table 4.

From the results shown for the first fifteen questions, the evidence indicates that classroom simulation training has no measurable effect on actual student teaching one year later. It is quite possible that the standards taught in classroom simulation training were of such a nature that a year's interim preparation also involved the same standards. Thus, no differences would be expected on a gross comparison of the experimental and control groups.

From evidence pertaining to question 16, there are very clear indications that classroom simulation training does have some effect on behavior. Students who underwent training were judged as being ready to assume full responsibility for a new class within the category of 1-3 weeks, and those who had no simulated experience were judged to be ready within 4-6 weeks.

Table 4.1 lists the proble recorded by supervising teachers as being most difficult for their student teachers (question 17 on the questionnaire). There are no clearly evident patterns of difference between the simulation and non-simulation groups.

Table 4

Summary of Comparisons Between the Simulated and Non-simulated Group on the Supervising Teacher Questionnaire

Question	Group	Mean	S.D.	t
1	Simulated	2.82	1.85	
	Non-simulated	2.30	1.06	1.00
2	Simulated	2.90	.70	
	Non-simulated	1.80	.63	0.97
3	Simulated	2.60	.70	
	Non-simulated	2.13	.64	1.39
4	Simulated	3.44	1.88	
	Non-simulated	3.11	1.76	0.38
5	Simulated	2.09	.70	
	Non-simulated	1.70	.95	1.06
6	Simulated	2.27	.79	
	Non-simulated	2.10	1.00	0.41
7	Simulated	1.36	.54	
	Non-simulated	1.50	.78	0.46
8	Simulated	2.18	.87	
	Non-simulated	1.80	.92	0.93
9	Simulated	1.82	.60	
	Non-simulated	2.50	1.18	1.64
10	Simulated	1.45	.46	
	Non-simulated	1.80	1.16	0.85
11	Simulated	2.45	1.73	
	Non-simulated	2.40	.93	0.09
12	Simulated	2.00	.63	
	Non-simulated	2.20	.18	0.25
13	Simulated	2.45	1.47	
	Non-simulated	2.00	1.22	0.84
14	Simulated	2.36	.70	
	Non-simulated	2.70	.46	1.00
15	Simulated	2.00	.40	
	Non-simulated	2.40	.27	1.50
16	Simulated	2.09	.63	
	Non-simulated	3.10	.89	2.52*

* $p < .02$

Table 4.1

Problems of Student Teachers as Recorded by Supervising Teacher

Problem	Simulation Group	Non-simulation Group
Fails to act quickly	0	1
Failure to plan for individual differences	3	2
Lack of voice control	4	3
Shyness	0	1
Fear	1	0
Lack of enthusiasm	1	0
Lack of discipline	2	1
Failure to encourage student initiative	0	2
Poor physical health	0	1
Lack of responsibility for actions	0	2
Difficulty in making decisions	0	1
Lack of leadership	0	1
Failure to attend to entire class	0	1
Lack of self confidence	0	3
Failure to make clear assignments	0	1
Difficulty in oral reading	1	0
Failure to acquire dictionary skills	1	0
Failure to relax at teaching	2	0
Failure to maintain or develop adequate student-teacher relations	3	0
Difficulty in planning and carrying out plans	1	1
Excessive talking	1	0
Failure to consider individual children as they relate to the group	0	1
Lack of dedication	1	0
Lack of sensitivity to children	1	0
Lack of humor	1	0

Discussion

The primary objective of the Follow-up Studies was to develop procedures for observing, recording, and evaluating the classroom behavior of Ts who have completed instruction in the simulated classroom. It was determined that, although not insurmountable, the problems involved in direct observation are much more difficult than were anticipated. Once the behavioral changes resulting from instruction in "Mr. Land's Sixth Grade" had been more clearly identified, and an appropriate means of recording Ts' classroom behavior had been developed, difficulty was encountered in establishing an acceptable level of observer agreement. Finally, it was determined that the specific occasions for observing the desired T-behavior occurred so irregularly and unpredictably in the normal classroom setting, that a large sample of T's classroom behavior would have to be recorded to provide sufficient data for evaluating transfer. Also, it was evident from the start that T's behavior in the classroom should be observed as soon as possible after his experience in the simulated classroom. There was a time-lapse of one year or more before Ts in the present study could be observed, during which they completed other "methods" courses in their Teacher Education program.

Vlcek (1965) observed Ts in the classroom during the college term following completion of simulation training and apparently had little difficulty observing Ts' reactions to problem incidents judged to be similar to those included in the simulated classroom materials. It is not clear in his report how long his trained observers recorded the classroom behavior of each T, but each T's classroom supervisor used the same observational procedure for a total of 2½ hours during one week (½ hour daily). By comparison, Es in the present study observed each T for a total of 60-90 minutes (20-30 minutes on each of three occasions). Another possible reason for the relatively successful procedure employed by Vlcek is that his observers were trained to identify and to evaluate only those problem situations which corresponded to the simulated problems. Es in the present study were trained to record verbal interactions between T and his pupils (and some of T's movements) at three-second intervals, regardless of the circumstances. The recording procedure employed in the present study is precise and objective, but the data are difficult to reduce by hand-processing techniques (i.e., prescribed interaction patterns are difficult to identify from among the many recorded events). Machine scoring techniques are being considered for use in the ongoing research effort.

Despite the difficulties encountered in data gathering, findings from both Vlcek's study and the present pilot effort indicate that there is transfer of learning from the simulated to the actual classroom. Vlcek's observational data indicate that in the simulated classroom, Ts learn principles of teaching which they employ in practice teaching soon after completion of their simulated experience; and questionnaire data from the present study suggests that, even after a year or more, Ts who have had simulation training are capable of assuming full responsibility for their first class weeks earlier than Ts who have not.

VIII The Ongoing Research Effort

The studies reported herein have already stimulated new research and developmental efforts. During the last year of the present research effort, after the preliminary results were known, two new lines of research were started: (1) studies aimed at increasing the efficiency and effectiveness of instruction in the simulated classroom, and (2) explorations into entirely new and different ways of using the classroom simulation materials. Included in the first line of effort are studies of prompting techniques (initiated by Dr. Paul Twelker) and of "instructional crowding," i.e., the attainment of two or more instructional objectives at once (initiated by the present writer and Dr. Twelker).

The second line of work is producing new applications of the original classroom simulation medium in counselor training and dental education, as well as teacher education. For example, Dr. James Beaird has recently sponsored pilot research in simulated interviews with a "client" who can lead the student counselor into several different paths of communication, depending on the questions asked. In the area of dental education, Mr. Victor Lund has developed a set of simulated "dental emergencies" which occur in a hypothetical dentist's office.

The one project which developed most directly from previous research findings, however, is a curriculum development effort aimed at the production of "low cost" classroom simulation materials. The new simulation materials are being designed for use either with groups or individuals, and so that they may be adapted for use by T with a minimum of direct supervision. Dr. Carl Wallen, Oregon State University, recently joined the project staff as associate to the writer on the project. Of most immediate interest is the preparation of a set of simulation materials for developing instructional skills identified with "discovery learning" at the intermediate grade level.

References

Adams, J. A. Some considerations in the design and use of dynamic flight simulators. In H. Geutzkow (Ed.), Simulation in social science: Read'ngs. N. J.: Prentice-Hall, Inc., 1962.

Chapman, R. L., Kennedy, J. L., Newell, A., & Biel, W. C. The System Research Laboratory's Air Defense Experiments, Mgmt Sci., 1959, 5 (3), 250-269.

Cunningham, L. L., News and comment, Elementary Sch. J., 1959, 60, 4-7

Flanders, N. A. Interaction analysis in the classroom: A manual for observers. (Revised edition). University of Michigan, 1964.

Frederiksen, N. In-basket tests and factors in administrative performance. In H. Guetzkow (Ed.), Simulation in social science: Readings. N. J.: Prentice-Hall, Inc., 1962.

Greenlaw, P. S., Herron, L. W., Rawdon, R. H. Business simulation in industrial and university education. N. J.: Prentice-Hall, Inc., 1962.

Guetzkow, H. Simulation in social science: Readings. N. J.: Prentice-Hall, Inc., 1962.

Guetzkow, H., Alger, C. F., Brody, R. A., Noel, R. C., & Snyder, R. C. Simulation in international relations: Developments for research and teaching. N. J.: Prentice-Hall, Inc., 1963.

Kersh, B. Y. Classroom simulation: A new dimension in teacher education. Teaching Research Division: Title VII, NDEA, Project no. 886, 1963.

Lund, Victor E. Teaching dental emergencies through simulation techniques. Final Report. Contracted Project #Ph 108-64-77 (P). Department of Health, Education and Welfare. Public Health Service.

Ryans, D. G. An information-system approach to theory of instruction with special reference to the teacher. SP-1079, System Development Corporation, 1963.

SAGE Training Plans Group, SAGE System Training Program. TM-250, System Development Corporation, 1957.

Schalock, Henry D., Beaird, James A., & Simmons, Helen. Motion pictures as test stimuli: An application of new media to the prediction of Complex Behavior. Final Report, Title VII Project no. 971. U. S. Office of Education, 1964.

University Council for Educational Administration, Simulation in administrative training.

Vlcek, C. W. Assessing the effect and transfer value of a classroom simulator technique. Michigan State University: Title VII, NDEA, Grant no. 7-32-0410-234 (Ed.D. dissertation project), 1965.

Winer, B. J. Statistical principles in experimental design. N. Y.: McGraw-Hill Book Co., Inc., 1962.

The appendix to this report is bound separately.

See the following document:

APPENDICES TO FINAL REPORTS

Classroom Simulation: Further Studies on Dimensions of Realism
Title VII Project Number 5-0848
Bert Y. Kersh, Principal Investigator

Prompting as an Instructional Variable in Classroom Simulation
Title VII Project Number 5-0950
Paul A. Twelker, Principal Investigator