The effects of trained and untrained proctors on student performance and satisfaction in a PSI course: a thesis...

Georganne White-Blackburn

University of the Pacific

Follow this and additional works at: https://scholarlycommons.pacific.edu/uop_etds

Part of the Education Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

THE EFFECTS OF TRAINED AND UNTRAINED
PROCTORS ON STUDENT PERFORMANCE AND
SATISFACTION IN A PSI COURSE

A Thesis
Presented to
the Graduates Faculty of
the University of the Pacific

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Georganne White-Blackburn
June 1977
This thesis, written and submitted by

GEORGANNE WHITE-BLACKBURN

is approved for recommendation to the Committee on Graduate Studies, University of the Pacific.

Department Chairman or Dean:

[Signature]

Thesis Committee:

[Signature]
[Signature]
[Signature]

Dated 5/31/77
ACKNOWLEDGMENTS

Of all of the people at the University of the Pacific who provided the author with advice, Dr. John R. Lutzker merits special thanks for his support and assistance as advisor and thesis committee Chair. Dr. Suzanne B. Hanser made a fine contribution and Dr. John H. Mabry came to the rescue on numerous occasions. Thank you, Philip Hanser, for volunteering so much time and wisdom. And, T.C., thanks for being there when you were needed.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td></td>
</tr>
<tr>
<td>Preliminary Proctor Survey</td>
<td>9</td>
</tr>
<tr>
<td>Subjects, Setting, and Course Format</td>
<td>13</td>
</tr>
<tr>
<td>Proctor Selection and Training</td>
<td>16</td>
</tr>
<tr>
<td>Response Definitions and Data Collection Procedures</td>
<td>18</td>
</tr>
<tr>
<td>EXPERIMENT I</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>EXPERIMENT II</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>46</td>
</tr>
<tr>
<td>Results</td>
<td>51</td>
</tr>
<tr>
<td>Discussion</td>
<td>62</td>
</tr>
<tr>
<td>GENERAL DISCUSSION</td>
<td>66</td>
</tr>
<tr>
<td>REFERENCE NOTES</td>
<td>73</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>74</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>77</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>80</td>
</tr>
</tbody>
</table>

ii
ABSTRACT

Among the variety of components comprising the PSI package—originally described by Keller (1968)—is the use of proctors which permit immediate scoring, tutoring, and personal-social interactions. Recent research has indicated that immediate scoring may not be an essential proctor function. However, personal-social interactions and tutoring may affect student academic performance and consumer satisfaction. Two experiments were conducted to evaluate the effects of the personal-social aspect of proctoring. A counterbalanced reversal plus control group design was utilized to investigate the effects of: (a) trained and untrained proctors and (b) trained and untrained proctors who participated in a lottery which was based on student performance and satisfaction. Data were collected on proctor behaviors during grading and student academic performance and consumer satisfaction ratings. Although trained proctors and proctors with lottery contingencies seemed to engage in the target behaviors more frequently, the performance of students enrolled in the course did not seem to be differentially affected in terms of the dependent measures investigated.
A personalized system of instruction (PSI) was originally described by Keller (1968). He advocated a variety of components which comprise the PSI package such as student-pacing, unit-perfection (mastery criterion), using the written word in teacher-student communication, and the use of proctors which permit repeated testing, immediate scoring, tutoring, and personal-social interactions. Studies have investigated the effectiveness of the PSI package compared with other teaching methodologies. For example, Sheppard and MacDermot (1970) compared a group of students who received their course material in a PSI format to a group of students who received the same course material in a more conventional manner which included small group discussion meetings, optional lectures, and short paper assignments. The results showed that students in the PSI group scored significantly higher on the final examination and rated the course more positively than did students in the conventional group.

In addition to investigating the PSI package as a whole, studies have been conducted to investigate the effectiveness of a number of individual components. For example, Johnston and O'Neill (1973) studied the effects of
different minimum performance criteria on the academic performance of college students in a PSI class. The results indicated that performance criteria controlled performance to a high degree, and that teacher-defined performance criteria produced superior performance when compared to student-determined criteria. Thus, teacher-defined mastery criteria seem to be a relevant component in the success of a PSI course.

Although a number of PSI components have been analyzed, there are few investigations of the role and function of the proctor component. Proctors, according to Keller (1968), represent an important aspect of PSI in that they provide immediate feedback to students, and serve both tutorial and social functions. Gaynor (1975) suggested that the proctor-student relationship should be analyzed in terms of the overall effectiveness of the PSI format. However, Gaynor (1975) did not suggest a research strategy for investigating the proctor role.

The two dependent variables often addressed in research of the PSI package and in research of components of PSI have been student academic performance and student consumer satisfaction. Often students perform well
academically and express satisfaction with the PSI course. However, whether the satisfaction with the course is due to the proctor component (proctor behaviors and/or proctor functions), has not been demonstrated. Thus, the effects of the proctor duties (immediate feedback, tutoring, and personal-social interactions) might be investigated both in terms of academic performance and course satisfaction.

Farmer, Lachter, Blaustein, and Cole (1972) investigated the role of proctoring and found that it was not necessary for proctors to provide immediate feedback for 100 percent of the quizzes taken by the students. In fact, Farmer et al. (1972) suggested that intermittent proctoring may be advisable so that the number of proctors required to administer the course can be reduced. Although Farmer et al. (1972) presented academic performance data to support the intermittent proctoring procedure, no data were presented on student satisfaction. Thus, intermittent proctoring may be a useful alternative to continuous immediate feedback in terms of student academic performance. However, student satisfaction may be affected by different schedules of proctor feedback.
Blackburn, Semb, and Hopkins (Note 1) addressed the issue of immediate scoring of student quizzes by evaluating the effectiveness of a self-grading procedure in which immediate feedback was provided when students graded their own quizzes using an instructor-prepared answer-key. The Blackburn et al. (Note 1) studies indicated that self-grading could be a useful alternative to proctor-grading by allowing the students access to immediate feedback. Thus, the number of proctors required to administer the course could be reduced without any decrement in student academic performance. However, the proctors continued to tutor, to record student grades, and to re-score a minimum number of randomly selected items on the self-graded quizzes. The self-grading procedure seemed to be useful in maintaining student academic performance, and the students rated self-grading more highly than proctor-grading. Thus, the value of proctors in a PSI course may be in providing other functions than immediate scoring of quizzes.

The tutorial function of proctors during testing was investigated by Hursh, Wildgen, Minkin, Minkin, Sherman, and Wolf (1975) in a study which evaluated whether
the opportunity to discuss quiz items with proctors and, thus, potentially correct incorrect quiz answers was effective in altering quiz accuracy. Hursh et al. (1975) found that although discussion did not seem to increase initial accuracy, students selected proctor discussion when given the option of discussion or no discussion.

The social function of proctors has not been directly assessed; however, Blackburn, White-Blackburn, and Lutzker (Note 2) investigated the effects of different formats of proctor-conducted discussion groups in a PSI course. Although the results indicated that different formats for discussion groups did not produce differential effects in academic performance when compared to each other and to a no-discussion group condition, the results of a questionnaire indicated that discussion groups and proctor discussion leaders received high student satisfaction ratings and that the ratings were higher for proctors than for any other PSI component.

If proctors serve a social and tutorial function which may relate to student satisfaction and academic performance, then what are the important proctor behaviors? Sanchez Sosa and Semb (Note 4) developed and validated a proctor selection procedure for personalized instruction.
Instructors, proctors, and students identified the 10 most desired proctor behaviors from a general list of 24 potential behaviors. Of the 10 behaviors selected, three related to the immediate feedback aspect (e.g., "serious-careful grading"), one related to the tutorial function ("knowledge of material") and six related to the social function (e.g., "smiling and eye contact"). However, the authors suggested that further research be conducted to ascertain whether proctors who engage in the 10 most desired proctor behaviors differentially affect students enrolled in the PSI course in terms of academic performance and consumer satisfaction.

Weaver and Miller (1975) developed a proctor training procedure which included a training manual and rehearsal. They found that their procedure was effective in training proctors to specified behavioral criteria. However, the behaviors and the criteria were selected by the authors as "appropriate" and, as with Sanchez Sosa and Semb (Note 4) and Gaynor (1975), there was no demonstration of which proctor behaviors of those selected were functional in terms of the academic performance and satisfaction of the students enrolled in the PSI course.
Although it has been demonstrated that PSI can be a productive course format for university classes and seems to promote academic performance and course satisfaction, the proctor behaviors which may contribute to student achievement and satisfaction within the PSI format have not been analyzed. Additionally, it has been demonstrated that experimenter-selected proctor behaviors can be trained, but it has not been demonstrated that there is a relationship between trained proctors and student performance. Nor has it been demonstrated that experimenter-selected "appropriate" proctor behaviors can account for the effects of the proctor component in a PSI course.

The purpose of this study was to investigate whether trained proctors produce differential effects in student academic performance and course satisfaction ratings. A proctor training program was developed following the Weaver and Miller (1975) model. However, the behaviors which were trained were determined by: (a) ascertaining which proctors currently serving in a PSI course were rated as "excellent" by students enrolled in the course, and (b) evaluating the "excellent" proctors to determine which behaviors consistently occurred among the highly rated proctors.
Two experiments were conducted to investigate whether proctors trained to specified criteria for "appropriate" proctor behaviors and who engaged in those behaviors during proctoring affected student performance and satisfaction in a PSI course. The first experiment compared trained and untrained proctors. The second experiment investigated whether a lottery contingency placed on proctors promoted the acquisition of "appropriate" proctor behaviors in the untrained proctor group.
Method

Preliminary Proctor Survey

In order to determine which behaviors were considered "appropriate" proctor behaviors by students, 10 proctor-student interactions were recorded on video-tape and evaluated by 24 undergraduate students enrolled in a PSI course. The 10 items selected as most desirable proctor behaviors in a proctor selection procedure developed by Sanchez Sosa and Semb (Note 4) were included. In addition, three items were incorporated into the questionnaire which was used to rate the proctor performance on the video-tapes (See Appendix A). Each item was followed by a scale which ranged from 1 to 7 (least favorable rating to most favorable rating). The proctors appearing on the video-tapes were graduate students in psychology who received no training in proctoring but had served as proctors for at least one semester. The undergraduate students who evaluated the video-tapes had never served as proctors but had completed at least one course taught in a PSI format.

Although all 10 proctors who appeared on the video-tapes received high ratings (range: 5.67-6.63), the proctors who received the highest and lowest ratings were
re-assessed to determine any differential behaviors between the highest and lowest rated proctors. The most highly rated proctors engaged in the following behaviors: (a) social: smiling at student, calling student by name, beginning the interview with a personal comment not related to the quiz (e.g., "You really look sharp today."). telling the student the outcome of the quiz, and concluding the interview with a positive, complimentary or encouraging comment (e.g., "You really seem to understand this material!"). (b) grading: delivering immediate feedback on the correctness of each answer while grading, prompting correct answers by restating the question, giving an example, and wording the prompts in a personal manner (e.g., "Whenever I had trouble with this concept, I think of the example of..."

The proctor behaviors which were targeted during the preliminary proctor survey seemed to be consistent with the behaviors trained by other researchers and regarded as "appropriate" proctor behaviors. For example, Weaver and Miller (1975) trained proctors to engage in behaviors included in three reliably observable response classes: preparation behaviors, prompting behaviors, and praise behaviors. Robin and Cook (Note 3) trained proctors to engage in nine component behaviors derived
from a sequential task analysis of the quiz-evaluation function of proctoring: greeting behavior, clear feedback, praise, telling students to proceed, listening without interruption, clear pass-fail statements, closing comments, non-quiz related questions, and administrative behaviors. However, many of the proctor training studies have not assessed whether the trained behaviors generalize from the training situation to the proctoring setting. Additionally, the question of whether the trained behaviors have an affect on student performance and satisfaction has not been addressed.

In a recent study by Robin and Heselton (1977), proctors were trained to engage in social behavior, feedback, praise, and prompting. During a follow-up observation of the proctors in the proctoring situation, social behavior seemed to maintain while the three other response classes reverted to near-baseline levels. An examination of review test performance of students assigned to the proctors did not produce a significant difference in academic performance. However, the emphasis of this study was to compare two different proctor training techniques and the lack of significant difference in academic performance could be attributed to both techniques being
equally effective in training proctors.

Thus, on the basis of the preliminary proctor survey, a questionnaire was developed which included items relating to the social and grading behaviors observed in the highly rated proctors and thought to be "appropriate" proctor behaviors by other PSI investigators. Additionally, a behavioral code was developed describing the targeted proctor behaviors and a data collection sheet was developed for use during observations conducted while proctors were proctoring to assess generalization of proctor training.

The purpose of the preliminary proctor survey was to assess whether highly rated proctors engaged in the behaviors regarded as "appropriate" proctor behaviors by other researchers and to develop a behavioral code and data collection system. Since the focus of this research was to assess the effects of trained and untrained proctors on student performance and satisfaction, an important aspect was to develop a behavioral code which was consistent with previous research and a data collection system which assessed generalization of proctor training which constituted the independent variable.
Subjects, Setting, and Course Format

One hundred and twenty-two undergraduate students enrolled in an elective psychology course entitled, "Contemporary Issues in Psychology," served as subjects. Total enrollment in the course was 220 students.

The course was divided into four parts and each part was further sub-divided into three units. There were weekly quizzes covering each unit of course material and each course part was followed by a review test which covered the preceding three units of material. Quizzes were administered by proctors in classrooms reserved for testing which contained individual desks for students and a file cabinet containing student files. Review tests were administered in a large lecture hall.

The course format consisted of two weekly lectures plus a weekly reading assignment. Study questions for the reading assignments were distributed and all test items were taken from the study questions. The course was both student- and instructor-paced. Students were required to take review tests on specified dates. In order to be able to take a review test, students were required to achieve mastery on the preceding reading quizzes. However, students were permitted to take the
three reading quizzes in each course part at their own pace. Thus, the instructor-specified due dates for review tests constituted instructor-pacing while student-pacing occurred in terms of the three reading quizzes required between review tests.

Reading quizzes. All reading quiz items were taken from the study questions covering each weekly unit of assigned reading material. Reading quizzes consisted of 15 items: 10 true-false items which required the student to write a one-sentence justification of his/her selection of true or false, and five short-answer essay questions. Proctors graded reading quizzes in front of the student providing immediate feedback and the opportunity for students to correct incorrect and partially correct answers. Mastery criterion was 90% (13.5 items correct) and students had the opportunity to retake as many versions as necessary to reach mastery.

Review tests. After each course part a review test was administered which consisted of 40 multiple choice items: 20-reading items, 10 lecture items, and 10 "think" items. The reading items were taken from the study questions for the three units of reading material which directly preceded the review test. Two-thirds
(66.7%) of the reading items on the review test appeared on Version A of the weekly reading quizzes for each course part. Since there was no opportunity for proctoring over lecture and "think" items prior to the review test, only reading items on review tests were included as a dependent measure in this study.

Pre-and post-test. An item pool of 240 multiple choice items taken from the study questions covering all units of reading material equally was constructed. Fifty items were randomly selected for the pre-test and 50 items were randomly selected for the post-test. Items were selected using a random number table and sampled equally each unit of reading material. On the first day of class, students were administered the pre-test and the post-test served as a final examination. The pre-test was administered to assess the knowledge of the students regarding the readings prior to the beginning of the course.

Course Material

Assigned readings were selected from seven popular books on the topic of child rearing practices: Between Parent and Child, Ginott, Child Behavior, Ilg and Ames, Parent Effectiveness Training, Gordon, Parents Are Teachers,
Becker, Baby and Child Care, Spock, Walden Two, Skinner, and Child Psychology: A Behavioral Approach to Everyday Problems, McIntire. Weekly reading assignments were approximately 175 pages in length ranging from 150 to 200 pages and were accompanied by approximately 50 study questions.

Proctor Selection

Twelve proctors were selected from a pool of volunteers prior to the beginning of the semester. All 12 students had similar histories in terms of PSI: all had completed the same PSI course as students, none had acted as a proctor previously, and none had completed more than the one PSI course at the university.

Proctor Training

Of the 12 proctors with similar PSI histories, five were randomly selected to receive training while seven were randomly assigned to the no-training group. After random assignment, all proctors were pre-tested in terms of proctor behaviors. The pre-test consisted of each proctor role-playing with a student-confederate to simulate proctor-student interactions.
The role-playing was recorded on video-tape and each proctor was given the same quiz to grade which contained correct, partially correct, and incorrect answers, and an instructor-prepared answer key to use during the grading of the quiz. After the conclusion of the semester, all proctors were post-tested and video-taped using the same procedure. Video-tapes of proctor behaviors were evaluated using the behavioral code and data collection sheet developed through the preliminary proctor survey.

The five students randomly assigned to proctor training received a three-part training program: one part consisted of viewing a video-tape of a model of correct proctor behaviors; during the second part, proctors role-played the modeled behaviors; finally, the proctors were required to discriminate among a number of video-taped examples of incorrect, partially correct, and correct proctor behaviors.

The training program was divided into three major groups of behaviors which were presented on the video-tape, role-played, and presented on the video-tape in varying degrees of correctness. The major groups of
behaviors were: (a) greeting which consisted of the proctor smiling at the student, calling the student by name, and making a personal comment not related to the quiz, (b) scoring which consisted of delivering verbal feedback while grading each quiz item, prompting correct answers by re-stating the question, giving an example, and wording the prompts in a personal manner, and (c) concluding the interview which consisted of the proctor telling the student the outcome of the quiz and ending the interview with a positive, complimentary or encouraging comment. Proctor training lasted approximately 90 minutes and proctors were considered to be trained when they could individually perform the modeled behaviors and could discriminate among the video-taped examples of correct, partially correct, and incorrect behaviors.

Response Definitions and Data Collection Procedures

Only students and proctors who signed a consent form for research were included in the study.

Proctor behaviors. Observers were present during all proctoring sessions for all proctors. The observers were trained to collect data on proctor behaviors and
were advised to not interact with proctors or students. The observers were not informed about the experimental hypothesis, proctor training, and the difference among the groups of trained and untrained proctors. Data were collected on the following proctor behaviors using an occurrence/non-occurrence system:

**Manager smiled at student** was defined as whether the proctor appeared to look at the student and smile at the beginning of proctoring before scoring the quiz.

**Manager calls student by name** was defined as whether the proctor uses first name, nickname or entire name of student before beginning to score quiz.

**Manager make a personal comment to student** was defined as whether proctor comments to student on something not related to the class or the quiz before scoring the quiz.

**Feedback on correctness** was defined as whether the proctor comments on the correctness of the answer. This included comments on partial correctness ("This part of the answer is very good.") and on incorrect answers ("This was a good try, but it isn't what the answer key says it should be.").
Number of prompts was defined as a tally of the number of prompts given by the proctor during proctoring on each item. A prompt was defined as remarks made by the proctor to elaborate on, clarify, define or simplify the question, and remarks made by the proctor regarding the incorrect portions of the student's answer. Prompts were counted between initiation of the item and final award of points for the item. Prompts made after the final award of points for the item were not counted ("You get ½ on this item because you couldn't define...").

Re-state question was defined as whether the proctor made a comment in which he/she interpreted the quiz item for the student ("This means that the concept is supported by Becker.").

Give example was defined as whether the proctor supplied the student with an example of a concept or procedure or elaborated upon the item using an example ("Let's say that you want to stop a kid from crying. Extinction would be if you ignored him everytime he cried.").

Personal prompts was defined as one or more of the prompts containing a personal opinion or experience or example of either the proctor of the student. Observers
were instructed to listen for "I" or "YOU" statements ("I used to have trouble with this one..."

Manager tells student outcome of quiz was defined as whether the proctor informed the student of his/her total points earned and/or whether student passed or failed the quiz. If the student prompted the proctor ("What did I get?"), it was counted as "NO" on this measure. The proctor must initiate the information, not wait to be asked by the student for it to be recorded as "YES."

Manager concluded interview with positive, complimentary or encouraging comment was defined as whether the proctor commented on the student's performance on the quiz in a potentially reinforcing manner or commented on the student's progress in the course or his/her abilities ("You sure did a good job today," or "I know that you can pass the next version with no problems after reviewing the materials again."). Negative comments were recorded as "NO" ("You sure are having trouble in this class." or "If you studied, you would pass.").

Student behaviors. The following measures were taken on student academic performance and consumer ratings.
Review test scores were defined as the number of points earned on reading items on the review tests. Review tests were graded by proctors using instructor-prepared criteria and proctors were not permitted to grade the review tests for their own students.

Pre- and post-test scores were defined as the number of correct answers on reading items. This number was converted into a percentage by dividing it by the total number possible times 100. The pre-test was graded by the experimenter and the post-test was graded by the proctors.

Student consumer satisfaction ratings were defined as the outcome of a questionnaire which was administered with each review test. Items on the questionnaire pertained to proctor behaviors and opinions about the class as a whole. Each item was accompanied by a scale of 1 to 7 (least favorable to most favorable) and students were instructed to select the number which most accurately reflected their opinions (See Appendix A).

Reliability

Interobserver agreement was assessed on proctor behaviors twice during the semester when a second observer independently collected data on the occurrence/non-occurrence
of the specified proctor behaviors. The second observer did not inform the proctors or the primary observers that reliability measures were being taken. Interobserver reliability was calculated by the number of agreements divided by the number of agreements plus disagreements times 100. An agreement was defined as both observers scoring a specific proctor behavior as occurring or not occurring. Reliability for proctor behaviors was 90% with a range of 80% to 100%.

Interobserver agreement on student behaviors was assessed by a second observer independently re-grading 10% of the review tests and post-tests using the same instructor-prepared answer key as was used by the proctors during initial grading. The second observer was not informed about the score originally given by the proctors during initial grading of the items. An agreement was defined as both the proctor score and the re-grade being the same for the item. Item-by-item reliability was calculated by the number of agreements divided by the number of agreements plus disagreements times 100. Interobserver agreement was 92% on the scoring of the items.
No reliability measures were taken on the results of the student consumer satisfaction questionnaire, the video-tape recordings of the proctor behavior pre- and post-test, and the student pre-test.

**Experiment I**

The purpose of Experiment I was to evaluate whether trained and untrained proctors differentially affect student performance and satisfaction in a PSI course.

**Method**

**Subjects and Setting**

The students, course format and materials, and the proctoring setting and training were identical to that which was described above.

Proctoring occurred on Monday morning in the testing rooms. Students enrolled into proctoring times on the first day of class without being informed of who the proctors were for each time and the differences among the groups of proctors. The only restriction on student enrollments in proctoring times was that there was a maximum of four students per proctor per testing hour for all proctoring times.

After proctor selection and prior to proctor training, three proctors were randomly assigned to the training
group (T), three proctors were randomly assigned to the untrained group (U), and two proctors were randomly assigned to the untrained-matched group (M).

The three proctors in the training group (T) completed proctor training to criteria while the proctors in the untrained group (U) and the untrained-matched group (M) were separately given an hour lecture on the course format, materials, and administrative duties of a proctor with no discussion regarding proctor behaviors or training.

Experimental design. A counterbalanced reversal plus comparison group design was utilized. The design can be illustrated as follows:

<table>
<thead>
<tr>
<th>PARTS OF THE COURSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

**TESTING TIME #1**

<table>
<thead>
<tr>
<th>ROOM A</th>
<th>ROOM B</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>T</td>
</tr>
</tbody>
</table>

**TESTING TIME #2**

<table>
<thead>
<tr>
<th>ROOM A</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>M</td>
</tr>
</tbody>
</table>

T = Trained Proctors, U = Untrained Proctors, M = Untrained-Matched

The counterbalanced reversal component of the design illustrated during Testing Time #1 controls for the
possible sequencing effect and the possible effect of varying difficulty of course materials between course parts. Thus, the same students experienced both trained and untrained proctors. At the end of each course part, proctors exchanged room assignments so that counterbalancing could occur. The counterbalanced reversal aspect of the design was intended to measure whether trained proctors produced differential effects in academic performance and consumer ratings in the same group of students who experienced both trained and untrained proctors.

However, a possible interactive effect could be hypothesized regarding the counterbalanced groups in that untrained proctors could be shaped by students who experienced both trained and untrained proctors during the semester. Thus, the untrained-matched group (M) was formed to provide normative information regarding the performance of proctors and students in a more typical PSI format.

The importance of the untrained-matched group (M) was to provide information on the effects of passage of time and proctoring experience over time on the behavior of the untrained proctors without the possible
confounding variable of the untrained proctors being shaped by students who could compare them with trained proctors during the same semester. The importance of the counterbalanced reversal portion of the design was to provide information on the effects of trained (T) and untrained (U) proctors on the same group of students without the possible confounding variables associated with groups of students of potentially different academic and PSI histories and possible problems with differences in difficulty of course materials between parts of the course.

Experimental procedures. Data were collected on the proctor and student behaviors described previously. Proctors were observed during proctoring one student each week. For experimental purposes, the 10 proctor behaviors were converted into a percentage of "appropriate" proctor behaviors per course part by totaling the number of "appropriate" proctor behaviors observed divided by the number of "appropriate" plus "inappropriate" behaviors times 100. Thus, each proctor received scores derived from direct observation of the pre-test, each course part, and the post-test (See Appendix A for data collection sheet).
Proctor training occurred four times during the semester: before proctoring the first quiz, between the third and fourth quizzes, between the sixth and seventh quizzes, and between the ninth and tenth quizzes. Thus, before changing room assignments after each part of the course, the trained proctors were re-trained to criteria. As in the initial training, the re-training utilized the same video-tape recordings and role-playing requirements.

Forty students enrolled into Testing Time #1: 18 students in Room A and 22 students in Room B. During Testing Time #2, the untrained-matched group (M) proctored 18 students.
Results

Figure 1 presents the mean percentage of appropriate proctor behaviors for each proctor in the trained and untrained groups. The video-tape recorded pre- and post-tests were evaluated using the same data sheets employed during direct observation during proctoring.

Since all proctors displayed very low variability on the direct observation measures, the data points for the course parts represent the mean percentage of appropriate proctor behaviors for the three observations conducted during each course part.

Figure 2 presents the mean percentage of appropriate proctor behaviors for the untrained-matched group.

Figure 3 presents the mean number of points earned on review tests by students enrolled in the trained and untrained proctor groups. Additionally, standard deviations are presented for each mean.
Figure 1
Mean percent of appropriate proctor behaviors for trained and untrained proctors on the video-taped pre- and post-tests and during proctoring.
Figure 2
Mean percent of appropriate proctor behaviors for untrained-matched proctors on the video-taped pre- and post-tests and during proctoring.
Figure 3
Mean number of points earned on review tests by students proctored by trained and untrained proctors.
(Vertical bars illustrate ± 1 standard deviation)
Figure 4 presents the mean number of points earned on review tests by students enrolled in the untrained-matched group of proctors. Standard deviations are presented.

-----------------------------

Insert Fig. 4 about here

-----------------------------

Figure 5 presents the frequency distribution of review test scores for each review test for each group of students enrolled in the trained and untrained groups of proctors. The number of students who earned scores in each percentile is presented.

-----------------------------

Insert Fig. 5 about here

-----------------------------

Figure 6 presents the frequency distribution of review test scores for each review test for each group of students enrolled in the trained and untrained groups of proctors. Since most of the students earned scores within the 90 percentile, Figure 6 presents the frequency distribution of the actual scores showing that most of the students achieved near-perfect review test scores on all of the review tests administered during the semester.
Figure 4
Mean number of points earned on review tests by students proctored by the untrained-matched proctors. (Vertical bars illustrate ± 1 standard deviation)
Figure 5
The number of students proctored by trained and untrained proctors earning review test scores in each percentile on TRAINED and each review test. UNTRAINED

PERCENT OF POINTS EARNED
Figure 7 presents the frequency distribution of review test scores for each review test for students enrolled in the untrained-matched proctor group. Again, most of the student earned scores within the 90 percentile, achieving near-perfect review test scores for all four review tests administered.

Figure 8 presents the mean percentage of correct items on the pre- and post-tests for the students enrolled in the trained and untrained proctor groups versus the students enrolled in the untrained-matched group. Since the students enrolled in the trained and untrained proctor groups experienced both trained and untrained proctors during the semester, their scores could not be separated in terms of pre- and post-test performance. Thus, Figure 8 illustrates that students in all groups entered the course with similar entry level skills and completed the course with similar performances on the post-test.
The number of students proctored by trained and untrained proctors earning review test scores on each review test.
Figure 7

The number of students proctored by untrained-matched proctors earning review test scores on each review test.
Figure 8
Mean percent of correct items on pre- and post-tests earned by students proctored by trained and untrained proctors, and untrained-matched proctors.
Table 1 presents the mean rating for trained and untrained proctors for each part of the course as measured on the questionnaire which accompanied each review test. Overall mean ratings for the 10 items which pertained to proctors are presented as well as individual mean ratings for the three items on the questionnaire which addressed the three target proctor behaviors: social, prompting, and feedback. In all cases, the distribution was skewed.

Insert Table 1 about here

Figure 9 presents the modal response on the questionnaire for the 10 items which pertained to proctors. Thus, the frequency distribution presented in Figure 9 illustrates the modal response for each student enrolled in the trained and untrained proctor groups. The selection of mode as a statistic was made to best reflect overall student attitudes towards proctors as measured on the 10 proctor items on the questionnaire since the distribution was skewed in all cases and the data were ordinal in nature. Frequency
### TABLE 1
Mean rating and ranges for trained and untrained proctors for each part of the course

<table>
<thead>
<tr>
<th>PARTS OF THE COURSE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRAINED</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>5.1</td>
<td>4.4</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>(1-7)</td>
<td></td>
<td></td>
<td>(3-7)</td>
<td>(1-7)</td>
</tr>
<tr>
<td>Prompting</td>
<td>5.6</td>
<td>5.3</td>
<td>6.2</td>
<td>5.3</td>
</tr>
<tr>
<td>(1-7)</td>
<td></td>
<td>(3-7)</td>
<td>(3-7)</td>
<td>(1-7)</td>
</tr>
<tr>
<td>Feedback</td>
<td>5.3</td>
<td>5.1</td>
<td>5.8</td>
<td>4.9</td>
</tr>
<tr>
<td>(2-7)</td>
<td></td>
<td>(1-7)</td>
<td>(2-7)</td>
<td>(1-7)</td>
</tr>
<tr>
<td><strong>UNTRAINED</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>4.3</td>
<td>5.1</td>
<td>4.9</td>
<td>5.1</td>
</tr>
<tr>
<td>(1-7)</td>
<td></td>
<td>(1-7)</td>
<td>(1-7)</td>
<td>(2-7)</td>
</tr>
<tr>
<td>Prompting</td>
<td>5.3</td>
<td>5.6</td>
<td>5.7</td>
<td>5.8</td>
</tr>
<tr>
<td>(2-7)</td>
<td></td>
<td>(1-7)</td>
<td>(1-7)</td>
<td>(2-7)</td>
</tr>
<tr>
<td>Feedback</td>
<td>5.1</td>
<td>5.3</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>(1-7)</td>
<td></td>
<td>(2-7)</td>
<td>(1-7)</td>
<td>(1-7)</td>
</tr>
</tbody>
</table>
distributions for trained and untrained proctors and for the untrained-matched group for the individual questionnaire items which addressed the three target proctor behaviors (social, prompting, and feedback) are presented in Appendix B.

Kolmogorov-Smirnov two-sample test (Hollander and Wolf, 1973) to detect differences in the underlying distribution of the modal data were performed in each of the review test situations. The null hypothesis of no difference in the distributions fails to be rejected at the .05 significance level for all review test situations.

Discussion

The results of Experiment I indicate that the trained proctors exhibited a higher rate of the target proctor behaviors than did the untrained and the untrained-matched proctors. Trained proctors performed the "appropriate" proctor behaviors approximately 30% of the time on the pre-test as compared with 77% of the time on the post-test and 76% of the time during direct observation of proctoring. The untrained proctors
Figure 9

The modal response on the questionnaire made by the number of students proctored by trained and untrained proctors for each review test.
and the untrained-matched proctors performed similarly to the trained proctors on the video-tape recorded pre-test ($U = 27\%$, $M = 20\%$). However, the untrained and untrained-matched proctors continued to exhibit low rates of "appropriate" proctor behaviors on the post-test and during direct observation of proctoring. Thus, proctor training increased the percentage of "appropriate" proctor behaviors exhibited by the trained proctors and the increase maintained over time and generalized to the proctoring situation.

Although the trained proctors exhibited a substantially higher rate of "appropriate" proctor behaviors during proctoring, there was an unsubstantial effect on student behaviors. Review test scores were similar in terms of mean score and standard deviations among the three groups of students enrolled in the different proctor groups: trained, untrained, and untrained-matched. Thus, the review test scores did not reflect the effects of proctor training.

Additionally, the results of the consumer satisfaction questionnaire did not reflect the observed differences among the three groups of proctors. In fact, the
frequency distributions for modal student responses and for individual questionnaire items were very similar among the three groups.

The results indicate that proctor training was effective in increasing the "appropriate" performance among the trained proctors. Additionally, untrained and untrained-matched proctors did not acquire the target proctor behaviors as a function of passage of time and proctoring experience. The results further indicate that student performance and satisfaction as measured in Experiment I were not affected by the observed differences among the groups of proctors.
Experiment II

The purpose of Experiment II was to evaluate whether untrained proctors acquired appropriate proctoring skills as a function of a contingency which permitted the proctors to participate in a lottery for money contingent upon adequate academic performance and high consumer ratings in the students enrolled in their proctoring time.

Method

Subjects and Setting

The students, course format and materials, and the proctoring setting and training were identical to that which was described above.

Proctoring occurred on Monday evening in the testing rooms. Students enrolled into proctoring times on the first day of class without being informed of who the proctors were for each time and the differences among the groups of proctors. As in Experiment I, the only restriction on student enrollments in proctoring times was that there was a maximum of four students per proctor per testing hour for all proctoring times.

After proctor selection and prior to proctor training, two proctors were randomly assigned to the training group (T) and two proctors were randomly assigned to the
untrained group (U). The untrained-matched group in Experiment II consisted of the same two proctors who had been randomly assigned to this group in Experiment I. Thus, different groups of trained and untrained proctors served as subjects in the two experiments while the same group of untrained-matched proctors served as a comparison group in both experiments.

The two proctors in the training group (T) completed proctor training to criteria while the proctors in the untrained group (U) and the untrained-matched group (M) were separately given an hour lecture on the course format, materials, and administrative duties of a proctor with no discussion regarding proctor behaviors or training.

Experimental design. A counterbalanced reversal plus comparison group design was utilized. The design can be illustrated as follows:
Experimental procedures. Data were collected on the proctor and student behaviors described previously. Proctors were observed during proctoring one student each week. For experimental purposes, the 10 proctor behaviors were converted into a percentage of "appropriate" proctor behaviors per course part by totaling the number of "appropriate" proctor behaviors observed divided by the number of "appropriate" plus "inappropriate" behaviors times 100. Thus, each proctor received scores derived from direct observation of the pre-test, each course part, and the post-test (See Appendix A for data collection sheet).

Proctor training occurred four times during the semester: before proctoring the first quiz, between
the third and fourth quizzes, between the sixth and seventh quizzes, and between the ninth and tenth quizzes. Thus, before changing room assignments after each part of the course, the trained proctors were re-trained to criteria. As in the initial training, the re-training utilized the same video-tape recordings and role-playing requirements.

Twenty-four students enrolled into Testing Time #3: 13 in Room A and 11 students in Room B. During Testing Time #2, the untrained-matched group proctored 18 students. Although the untrained-matched group of proctors and students were the same for Experiments I and II, the students and proctors who served as subjects in the trained and untrained groups in Experiments I and II were separate with no overlap.

Trained and untrained proctors in Experiment II were informed that they would have the opportunity to participate in a lottery for $50.00 at the end of the semester contingent upon the adequate academic performance of their students coupled with high ratings on the satisfaction questionnaire which accompanied each review test. Adequate
academic performance was defined as all students earning at least 70% on the review test. Proctors were deliberately not given a definition of "high ratings." However, for experimental purposes in determining whether proctors qualified for the lottery, high ratings were defined as a mean rating on the 10 target items which was equal to or greater than four on a scale of 1 to 7 (least favorable to most favorable).

Proctors were told that they could earn one lottery ticket for each of the four review tests in which their students academically performed to criterion and gave the proctors a "high rating." However, proctors were not given feedback regarding the performance of their students or the results of the consumer ratings until the time of the lottery at the end of the semester. Proctor behaviors were observed to determine whether the contingencies were effective in promoting the acquisition of "appropriate" proctor behaviors in the untrained group. Student behaviors were observed to assess whether proctor behavior differentially affected student behaviors.

At the time of the lottery, it was found that all of the proctors serving during Testing Time #3 qualified
for the lottery on all four review tests. Student academic performance and ratings were sufficient to permit both trained and untrained proctors to qualify for the lottery with the maximum number of tickets possible.

Results

Figure 10 presents the mean percentage of appropriate proctor behaviors for each proctor in the trained and untrained groups. The video-tape recorded pre- and post-tests were evaluated using the same data sheets employed during direct observation during proctoring.

Insert Fig. 10 about here

Since all proctors displayed very low variability on the direct observation measures, the data points for the course parts represent the mean percentage of appropriate proctor behaviors for the three observations conducted during each course part.

Figure 11 presents the mean percentage of appropriate proctor behaviors for the untrained-matched group.

Insert Fig. 11 about here
Figure 10
Mean percent of appropriate proctor behaviors for trained and untrained proctors on the video-taped pre- and post-tests and during proctoring.
Figure 11
Mean percent of appropriate proctor behaviors for untrained-matched proctors on the video-taped pre- and post-tests and during proctoring.
Figure 12 presents the mean number of points earned on review tests by students enrolled in the trained and untrained proctor groups. Additionally, standard deviations are presented for each mean.

--- Insert Fig. 12 about here ---

Figure 13 presents the frequency distribution of review test scores for each review test for each group of students enrolled in the trained and untrained groups of proctors. Since most of the students earned scores within the 90 percentile, Figure 13 presents the frequency distribution of the actual scores showing that most of the students achieved near-perfect review test scores on all of the review tests administered during the semester.

--- Insert Fig. 13 about here ---

Figure 14 presents the mean percentage of correct items on the pre- and post-tests for the students enrolled in the trained and untrained proctor groups versus the students enrolled in the untrained-matched group. Since the students enrolled in the trained and untrained proctor groups experienced both trained and untrained proctors.
Figure 12

Mean number of points earned on review tests by students proctored by trained and untrained proctors.

(Vertical bars illustrate ± 1 standard deviation)
Figure 13
The number of students proctored by trained and untrained proctors earning review test scores on each review test.
during the semester, their scores could not be separated in terms of pre- and post-test performance. Thus, Figure 14 illustrates that students in all groups entered the course with similar entry level skills and completed the course with similar performance on the post-test.

Table 2 presents the mean rating for trained and untrained proctors for each part of the course as measured on the questionnaire which accompanied each review test. Overall mean ratings for the 10 items which pertained to proctors are presented as well as individual mean ratings for the three items on the questionnaire which addressed the three target proctor behaviors: social, prompting, and feedback. In all cases, the distribution was skewed.

Figure 15 presents the modal response on the questionnaire for the 10 items which pertained to proctors.
Figure 14

Mean percent of correct items on pre- and post-test earned by students proctored by trained and untrained proctors, and untrained-matched proctors.
### TABLE 2

Mean rating and ranges for trained and untrained proctors for each part of the course

<table>
<thead>
<tr>
<th>PARTS OF THE COURSE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAINED</td>
<td>5.3</td>
<td>5.1</td>
<td>5.2</td>
<td>5.4</td>
</tr>
<tr>
<td>UNTRAINED</td>
<td>5.0</td>
<td>5.3</td>
<td>5.3</td>
<td>4.7</td>
</tr>
<tr>
<td>TRAINED</td>
<td>6.0</td>
<td>5.5</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Social</td>
<td>(2-7)</td>
<td>(1-7)</td>
<td>(2-7)</td>
<td>(2-7)</td>
</tr>
<tr>
<td>Prompting</td>
<td>6.2</td>
<td>5.6</td>
<td>6.3</td>
<td>6.4</td>
</tr>
<tr>
<td>Feedback</td>
<td>6.2</td>
<td>5.6</td>
<td>6.4</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>(1-7)</td>
<td>(1-7)</td>
<td>(1-7)</td>
<td>(3-7)</td>
</tr>
<tr>
<td>UNTRAINED</td>
<td>5.5</td>
<td>6.0</td>
<td>5.5</td>
<td>5.3</td>
</tr>
<tr>
<td>Social</td>
<td>(1-7)</td>
<td>(2-7)</td>
<td>(2-7)</td>
<td>(3-7)</td>
</tr>
<tr>
<td>Prompting</td>
<td>5.5</td>
<td>6.2</td>
<td>6.1</td>
<td>5.8</td>
</tr>
<tr>
<td>Feedback</td>
<td>5.5</td>
<td>6.2</td>
<td>5.8</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>(1-7)</td>
<td>(1-7)</td>
<td>(2-7)</td>
<td>(3-7)</td>
</tr>
</tbody>
</table>
Thus, the frequency distribution presented in Figure 15 illustrates the modal response for each student enrolled in the trained and untrained proctor groups. The selection of mode as a statistic was made to best reflect overall student attitudes towards proctors as measured on the 10 proctor items on the questionnaire since the distributions are skewed in all cases and the data are ordinal in nature. Frequency distributions for trained and untrained proctors for the individual questionnaire items which addressed the three target proctor behaviors (social, prompting, and feedback) are presented in Appendix B.

Kolmogorov-Smirnov two-sample test (Hollander and Wolf, 1973) to detect differences in the underlying distribution of the modal data were performed for each review test. The null hypothesis of no difference in the distributions fails to be rejected at the .05 significance level for all review test situations.
The modal response on the questionnaire made by the number of students proctored by trained and untrained proctors for each review test.
Discussion

The results of Experiment II indicate that proctor training was effective in increasing "appropriate" performance among the trained proctors. The effects of the proctor training program were replicated in that the trained proctors increased from a pre-test of approximately 25% "appropriate" behaviors to a post-test of approximately 81%. Additionally, proctor training generalized to the proctoring situation and the high rate of proctor behaviors among the trained proctors maintained over time. Thus, trained proctors in Experiment II exhibited a similar rate of target proctor behaviors to the trained proctors in Experiment I.

The effects of the lottery contingency appear in the observed behaviors of the untrained group when compared to the untrained-matched group of proctors. Although untrained and untrained-matched proctors exhibited low rates of target behaviors on the pre-test ($U = 25\%, \overline{M} = 20\%$), post-test scores indicate that the untrained proctors exhibited approximately 45% "appropriate" behaviors while the untrained-matched proctors engaged in the target behaviors approximately 26% of the time. Thus, the lottery contingency increased the rate of observed proctor behaviors among
the untrained proctors when compared to the performance of the untrained-matched group and to the performance of the untrained proctors in Experiment I.

The effects of the increased "appropriate" proctor behaviors was not apparent in terms of student academic performance and satisfaction, however. The mean and standard deviation on review test scores was similar for all three groups of students enrolled in the different proctor groups: trained, untrained, and untrained-matched. Additionally, the consumer satisfaction questionnaire did not reflect observed differences among the three proctor groups. In fact, the frequency distributions for modal student responses and for individual questionnaire items were very similar among the three groups.

The results indicate that proctor training was effective in increasing "appropriate" proctor performance among the trained proctors. Additionally, the lottery contingency increased untrained proctor performance on the observed behaviors. However, the effects of increased performance of the target behaviors were not reflected in the student academic performance measures
and the consumer satisfaction questionnaire.

Proctor training versus proctor contingency could be evaluated in terms of a cost-effectiveness model. The proctor training program was more costly than the contingency since the training included preparation of video-tapes and training materials as well as 90 minutes to administer four times during the semester; while the contingency cost a total of $50.00 plus approximately 30 minutes to review student performance and satisfaction which qualified proctors for the lottery. Since the lottery contingency increased observed proctor behaviors among the untrained proctors, the contingency could be a useful alternative to proctor training in terms of cost.

The effectiveness of the contingency was indicated by comparing the observed behaviors of the trained, untrained, and untrained-matched groups of proctors. While the trained proctors exhibited the target behaviors approximately 82% of the time during direct observation, the untrained scored approximately 63% while the untrained-matched scored approximately 38%. Thus, the contingency served to increase target behaviors among untrained proctors so that they performed better than the untrained-
matched but not as well as the trained proctors.

Although the cost of proctor training was higher than the contingency and the contingency was effective in increasing the observed behavior of the untrained proctors, the ultimate effectiveness decision should be based on the effects of the observed behaviors on the student performance and satisfaction measures. Student performed well and were satisfied with both trained and untrained proctors. Additionally, students were satisfied with and performed well with untrained-matched proctors. Thus, from a cost-effectiveness perspective, Experiment II indicated that untrained-matched proctors were no less effective in terms of student performance and satisfaction although they exhibited fewer of the target proctor behaviors.
General Discussion

The proctor training program increased the rate of the target proctor behaviors among the trained proctors in Experiments I and II. Additionally, the trained proctors maintained the increased target behaviors over time and the effects of training generalized to the proctoring situation.

Generalization of proctor behaviors may be due, in part, to the reactivity of observers present during all proctoring sessions. Although previous research (Robin and Heselton, 1977) indicated that generalization of target behaviors from proctor training to actual proctoring may not occur in all cases, the generalization indicated in this study may be due to the implicit contingencies surrounding the presence of observers in the proctoring situation. Thus, further research may result in the development of strategies which prompt generalization of proctor training.

Although trained proctors exhibited the trained target behaviors during proctoring, there was an unsubstantial effect of different proctor groups on student performance and satisfaction in the course. The apparent lack of effect on student behaviors could
be attributed to the assumption that the PSI format utilized in this course maintained student performance and satisfaction in spite of differences among groups of proctors in terms of the observed proctor behaviors.

When a PSI class contains important components such as unit perfection (mastery criterion), student-pacing, and immediate feedback of quiz results, it seems that variations in proctor behaviors, quiz item formats, and ancillary discussion groups may not decrease student performance. White-Blackburn, Blackburn, and Lutzker (1976) demonstrated that different quiz item formats (objective vs. subjective) did not produce differential effects in student academic performance on review tests which contained subjective and objective items. Additionally, Blackburn, White-Blackburn, and Lutzker (Note 2) demonstrated that different formats of discussion groups did not differentially affect student academic performance. Thus, it could be assumed that the basic PSI model can be altered along some dimensions without any decrement in student academic performance.

Although it may be tempting to conclude that proctor training is not effective in increasing student
performance and satisfaction in PSI courses, this conclusion would be hasty based on the data presented in Experiments I and II. The lack of substantial effect of proctor training could be due to: (a) the PSI format utilized in the course which may have been sufficiently powerful to shadow the effects of the trained proctors, (b) the selection of dependent measures which may have not been sensitive to the effects of the trained proctors, (c) the choice of the target behaviors which may not be the essential components in maximally successful proctoring, and (d) the restricted range of student behaviors as measured by the dependent variables.

The course may have been sufficiently well programmed, the study questions sufficiently informational, and the PSI format utilized could have been effective enough to produce a high level of performance and satisfaction among all of the students. Thus, the role of proctors and the difference among proctor groups might have been a more powerful variable if other elements of the course were less adequate.

A ceiling effect in student academic performance is often a product of the PSI format. Thus, review test
scores may not be sufficiently sensitive to reflect
difference among proctors. Additionally, the satisfaction
questionnaire may not be sensitive to changes in student
opinion and may suffer from the same ceiling effect as
the review tests. That is, students may have been
pleased with the course, the instructor, the format,
and the readings resulting in a generally positive evalua-
tion which was not sensitive to the contribution of
proctor behaviors. Perhaps, a questionnaire administered
with each review test is not the best strategy for
assessing student satisfaction with the course.

The target behaviors which were selected as
"appropriate" proctor behaviors may not be the important
behaviors for successful proctoring. Although social
behavior, prompting, and feedback seem to be generally
acknowledged as important proctor behaviors, there may
be other behaviors which are more effective in controlling
student performance and satisfaction in a PSI course.
Further research is indicated in the selection of target
proctor behaviors and in the assessment of the effects
of a variety of proctor behaviors on dependent measures
regarding students.
The restricted range of student behaviors as measured by the dependent variables could account for the apparent lack of effect of proctor training. Since the course had never been taught before, the study questions, quiz items, and review test items had not been analyzed prior to this study. Students in all groups achieved relatively high pre-test scores which might indicate that the materials were not sufficiently difficult to permit discrimination among the contributions of trained and untrained proctors in terms of student academic performance. Although students read seven books, were required to reach mastery on 12 reading quizzes, and took four review tests and a final examination during the course, the quality of the reading materials and test items may have differed sufficiently to affect student performance independent of the PSI format and proctor training.

The restricted range of student responses on the questionnaire could be due to the circumstances surrounding the administration of the questionnaire. That is, students completed the questionnaire after completing a review test
rather than directly after being proctored, students completed the questionnaire in the lecture hall rather than in the testing room which may constitute sufficiently different stimulus conditions to affect student responses, and students were not informed whether the results of the questionnaire would affect proctors which may have prompted students to protect their proctors from instructor criticism rather than evaluate the proctor performance.

Thus, the results from Experiments I and II indicate that proctor training can be effective when coupled with generalization contingencies. However, whether trained proctors differentially affect student performance and satisfaction can not be answered due to the questionable merit of the dependent measures examined. Further research is indicated to develop dependent measures which reflect a variety of proctor behaviors.

The results from Experiment II indicate that contingencies may be a useful alternative to proctor training from a cost-effectiveness perspective. However, the ultimate decision regarding proctor training versus proctor contingencies lies in the evaluation of proctor
behaviors essential to the PSI format. It seems to be a moot point to discuss proctor training versus proctor contingencies until the specific proctor behaviors of importance are discovered. Then, further research can discover the relative effectiveness of proctor training and contingencies.

Although this study did not answer the question of whether trained and untrained proctors differentially affect student academic performance and satisfaction, the data seem to indicate that in order to measure the effects of different groups of proctors on student performance, different measures should be devised. Thus, the issue of cost-effectiveness may be important in that if the usual dependent variables typically utilized to evaluate the effectiveness of PSI as a whole and the components of PSI are not sensitive enough to show differences among the groups of proctors, then are the differences so small that their importance is questionable in term of PSI as a whole? Additionally, if the dependent measures typically used in PSI do not show substantial differences, then is it useful to expend time and money training proctors to engage in behaviors which produce a small effect in the overall effectiveness of the PSI format?
Reference Notes


References


APPENDIX A

Preliminary Proctor Survey Rating Form
Student Manager Data Sheet #1, Reading Quizzes
Student Consumer Satisfaction Questionnaire
PROCTOR RATING FORM

PROCTOR'S NAME:

1. Proctor comments on correct answers in the quiz.
   Never 1 2 3 4 5 6 7 Frequently
2. Proctor comments on partially correct answers in the quiz.
   Never 1 2 3 4 5 6 7 Frequently
3. Proctor is able to rephrase quiz questions.
   Very poorly 1 2 3 4 5 6 7 Very well
4. Clarity and adequacy of proctor's explanations.
   Very poor 1 2 3 4 5 6 7 Very good
5. Creativeness and novelty of proctor's examples.
   Very poor 1 2 3 4 5 6 7 Very good
6. Proctor's knowledge of material.
   Very poor 1 2 3 4 5 6 7 Very good
7. Proctor's general helpfulness.
   Very low 1 2 3 4 5 6 7 Very high
8. Did proctor praise the student's best efforts?
   Almost never 1 2 3 4 5 6 7 Almost always
   Very unfriendly 1 2 3 4 5 6 7 Very friendly
10. Did the proctor seem interested and concerned about the student?
    Not really 1 2 3 4 5 6 7 Yes
11. I have been proctored by this person more than once in Behavior Change I.
    Yes No
12. In terms of the proctor-student relationship, I know this proctor
    only from this tape 1 2 3 4 5 6 7 very well from Behavior Change I.
13. From my experience with this proctor, I think that his/her performance
    on the video-tape was not typical 1 2 3 4 5 6 7 very typical
PROCTOR BEHAVIOR OBSERVATION RECORDING SHEET

STUDENT MANAGER DATA SHEET #1  READING QUIZZES

MANAGER'S NAME: ___________________________  DATE: ___________  OBSERVER: ___________________________

Student began reading quiz: ___________________________  Ended quiz: ___________________________

INITIAL MANAGER-STUDENT INTERACTION FOR GRADING:

Grading of Reading Quiz began: ___________________________

A. Manager smiled at student.  YES  NO  ?
B. Manager called student by name.  YES  NO  ?
C. Manager made personal comment to student.  YES  NO  ?

<table>
<thead>
<tr>
<th>Quiz Item</th>
<th>Score</th>
<th>Feedback on Correctness</th>
<th># of Prompts</th>
<th>Re-State Question</th>
<th>Give Example</th>
<th>Personal?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. T/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Essay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Essay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Essay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Essay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Essay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D. Manager told student outcome of quiz.  YES  NO  ?
E. Manager concluded interview with positive, complimentary or encouraging comment.  YES  NO  ?
F. Grading of Reading Quiz ended: ___________________________

STUDENT'S NAME: ___________________________  READING QUIZ #: ___________  VERSION: ___________________________
STUDENT CONSUMER SATISFACTION QUESTIONNAIRE
COMMENTS ON THE COURSE

NAME: ___________________________________ TESTING DAY: _______ TESTING TIME: _______

INSTRUCTIONS: Please circle the number below each item which best reflects your opinion. This questionnaire is extremely important and your answers will be taken very seriously. So, please take the time to read each item carefully and give your honest opinion.

1. I feel that I am learning a great deal in this class.
   STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

2. The lectures are very exciting and interesting.
   STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

3. I like the readings assigned for this class very much.
   STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

4. My proctors usually comment on correct and partially correct answers on my reading quizzes.
   NEVER 1 2 3 4 5 6 7 FREQUENTLY

5. My proctors usually rephrase the questions for me if my answers aren't correct.
   NEVER 1 2 3 4 5 6 7 FREQUENTLY

6. My proctors usually have clear and adequate explanations for the quiz questions and the correct answers in the answer keys.
   NEVER 1 2 3 4 5 6 7 FREQUENTLY

7. When my proctors give an example, it is usually creative and novel.
   ABSOLUTELY 1 2 3 4 5 6 7 NEVER

8. I think that my proctors have an excellent knowledge of the course materials.
   STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

9. My proctors are very helpful.
   STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

10. My proctors praise my best efforts and are very reinforcing to me.
     STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

11. I feel that my proctors are very friendly.
     STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

12. In fact, my proctors seem very interested and concerned about each student.
     NOT REALLY 1 2 3 4 5 6 7 YES

13. I think that the whole idea of PSI (the way this class is run) and proctors is FANTASTIC !!!! 1 2 3 4 5 6 7 YUK!!

14. If I had my way, all of my classes would be taught just like this one.
     STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

15. I like this class so much that I may major in psychology.
     STRONGLY AGREE 1 2 3 4 5 6 7 STRONGLY DISAGREE

Any additional comment you may wish to make will be welcomed. Thank you for completing this questionnaire. Your comments will help to make this course better.
APPENDIX B

EXPERIMENT I: Frequency Distributions for Trained and Untrained Proctors

Proctor Social Behaviors
Proctor Feedback Behaviors
Proctor Prompting Behaviors

EXPERIMENT II: Frequency Distributions for Trained and Untrained Proctors

Proctor Social Behaviors
Proctor Feedback Behaviors
Proctor Prompting Behaviors

EXPERIMENTS I & II: Frequency Distributions for Untrained-Matched Proctors

Proctor Social Behaviors
Proctor Feedback Behaviors
Proctor Prompting Behaviors
EXPI

REVIEW TEST #1 | REVIEW TEST #2 | REVIEW TEST #3 | REVIEW TEST #4

10 -
9 -
8 -
7 -
6 -
5 -
4 -
3 -
2 -
1 -
0 -

NUMBER OF STUDENTS

1 2 3 4 5 6 7

Dr X = 5.1
Dr X = 4.4
Dr X = 5.2
Dr X = 5.1

Dr X = 4.3
Dr X = 5.1
Dr X = 4.9
Dr X = 5.1

PROCTOR "SOCIAL" ITEM ON QUESTIONNAIRE
(ITEM #12)

Trained proctor group
Untrained proctor group
EXP.I

NUMBER OF STUDENTS

FEEDBACK ITEM ON QUESTIONNAIRE (ITEM #4)
EXP. I

REVIEW TEST #1
- Y = 5.6
- X = 6.4

REVIEW TEST #2
- Y = 5.3
- X = 6.6

REVIEW TEST #3
- Y = 6.2
- X = 5.7

REVIEW TEST #4
- Y = 5.3
- X = 5.8

NUMBER OF STUDENTS

- TRAINED PROCTOR GROUP
- UNTRAINED PROCTOR GROUP

PROCTOR "PROMPTING" ITEM ON QUESTIONNAIRE
(ITEM #5)
EXP. II

Review Test #1

\[ \bar{x} = 6.2 \]

\[ s = 5.5 \]

Review Test #2

\[ \bar{x} = 5.6 \]

\[ s = 6.2 \]

Review Test #3

\[ \bar{x} = 6.3 \]

\[ s = 6.1 \]

Review Test #4

\[ \bar{x} = 6.4 \]

\[ s = 5.8 \]

PROCTOR "PROMPTING" ITEM ON QUESTIONNAIRE

(ITEM #5)