

INFORMATION TO USERS

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

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

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again -- beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

MASTERS THESIS

M-7489

SELDEN, Paul Hubert, III

REDUCED REMEDIATION IN A PERSONALIZED SYSTEM
OF INSTRUCTION (PSI) STYLED PSYCHOLOGY COURSE
AS A FUNCTION OF PREQUIZ DISCUSSION OF STUDY
QUESTIONS.

Western Michigan University, M.A., 1975
Psychology, experimental

Xerox University Microfilms, Ann Arbor, Michigan 48106

REDUCED REMEDIATION IN A
PERSONALIZED SYSTEM OF INSTRUCTION (PSI)
STYLED PSYCHOLOGY COURSE AS A FUNCTION OF
PREQUIZ DISCUSSION OF STUDY QUESTIONS

by

Paul Hubert Selden III

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment
of the
Degree of Master of Arts

Western Michigan University
Kalamazoo, Michigan
August 1975

ACKNOWLEDGEMENTS

I wish to express my heartfelt appreciation to all those who have made my education at Western Michigan University possible. I give thanks to Drs. Daniel E. Hursh, Paul T. Mountjoy, and Galen J. Alessi for their assistance, counsel, and encouragement during the course of this experiment. In particular, Dr. Hursh deserves credit for the many hours of patient constructive criticism I have received from him, thus intimately guiding and facilitating the course of this research. My thanks go to Bev Polay, Andrea Johnson, Carl Kalitta, Leo McNamee III, and Bruce Reed for their assistance in running the day to day operations of this experiment, and I thank the many students who participated in this research. Several persons not standing in the foreground of this experiment have played an important role in my education at this university nevertheless. Therefore I thank Dr. James Ulness, who first "showed me the ways" of operant conditioning and behaviorism. Dr. David O. Lyon and Dr. Mountjoy deserve my gratitude for favoring me with their friendly advice and patronly spirits, greatly enhancing my educational experience and easing the financial burdens of a graduate education. Last, but certainly not least, I must thank my dear father, the late Cmdr. Paul H. Selden, Jr., and my wonderful mother, Dr. Margery S. Selden, whose biological, moral, and economic support made my attendance at this fine institution possible in the first place.

Paul H. Selden

TABLE OF CONTENTS

	Page
INTRODUCTION	1
METHOD	7
Students	7
Settings and Sessions	8
Procedures	8
Experimental Design	10
Measurement and Reliability	17
RESULTS	18
DISCUSSION	23
FIGURE LEGENDS AND FIGURES	34
APPENDIX A	39
APPENDIX B	41
REFERENCES	43

INTRODUCTION

Since its inception Personalized Systems of Instruction (PSI) has offered itself as an alternative to more traditional methods of education (Keller, 1966). Many of the early proponents of PSI have submitted their methods to experimental or partial experimental scrutiny, noting that in contrast to the more common lecture-assignment-infrequent testing format, PSI educated students have tended to receive more "As" and "Bs", and fewer "Cs", "Ds", and "Fs", score higher on final exams, and indicate more "satisfaction" with the PSI format, as well as perform better on follow up "retention" tests (Keller, 1968; Sheppard and MacDermot, 1970; McMichael and Corey, 1969; Corey, McMichael and Tremont, 1970; Corey, Valente and Shamow, 1971).

The PSI methodology has not been restricted to the fields of social science; it has been successfully adapted to the fields of physics (Green, 1971), mechanical engineering (Hoberock, 1971), engineering (Koen, 1970), biology (Woore, Mahan and Ritts, 1969), and statistics (Myers, 1970), among many others (PSI Newsletter, 1971-).

More recently the components comprising the PSI package have been experimentally examined to determine their efficacy and consequent necessity within the program. Gallegos (1968), Lloyd (1971), and Miller, Weaver and Semb (1974) agree that students tend to procrastinate if allowed to freely pace their own progress within the course, but that instructor imposed deadlines and contingencies effectively insure that students will maintain progress satisfactorily. Semb (1974) found that short units of study help students get higher grades than when they

are given longer ones. Dalis (1970), Gustafson and Toole (1970), and Semb, Hopkins and Hursh (1973) agree that when study questions are provided covering a given unit's material student exam performance is raised. Semb (1974) and Dustin (1971) have found that frequent testing likewise facilitates student performance, as compared with relatively infrequent testing. Johnston and O'Neill (1973) and Semb et al. (1973) agree that when high grades are contingent upon high performance ("mastery") students tend to perform better. In his thorough review of experimental analyses of PSI Hursh (unpublished) notes that the use of proctors as a necessary part of PSI has not been established, and that typical lectures when used as "optional motivators do not function as reinforcers and are not essential to the maintenance of high levels of exam performance ..." (p. ii).

Others have modified the format of the original (Keller, 1966 and 1968) PSI approach, changing certain portions of it to discover sound alternatives. Ferster (1968), Johnston and Pennypacker (1971) and Alba and Pennypacker (1972) have successfully employed oral interviews instead of written tests. The latter two studies used a measure of rate correct instead of using a percent correct measure. Sherman (1971) and Gaynor and Wolking (1974) have successfully recruited proctors from students currently enrolled in the course rather than from students who had finished the course previously. The present experiment was designed to examine yet another variation, based on the following rationale.

In many PSI structured courses the student preparing for his

weekly unit test has available study objectives (often phrased in the form of questions) which must be mastered (or answered), and some sort of material to be studied, from which the answers to the study questions may be obtained. The student answers the study questions in some fashion, then must somehow prepare to respond (on the exam, often to those same study questions) in the absence of all materials save the exam questions themselves. To accomplish this a good student may spend time practicing, often covering the answer he has written to the objective or study questions, attempting to answer it, looking again at the answer, covering it again, repeating the process until he is able to indeed emit the answer in the presence of the test question alone. Of course, should the answer he is learning be wrong in the first place, the answer he gives during the test situation will be wrong, too. Now, if the student does not accurately answer the actual exam questions during the first test over the unit, the PSI format provides for repeated attempts, at no academic penalty to the student. Unless the student requires many remedial tests over many units, the only real penalty involves a repetition of the study process in order to prepare for remedial testing. In the studies cited above most of the students do pass after repeated testing, so, for one reason or another the student eventually masters the material and receives a high grade.

Mentioned above was one "hidden cost" associated with the PSI provision for remedial testing, namely, that the student must spend time restudying material, he may have "wasted" time taking the test which was failed, must expend more time and energy coming to class again

and taking the test again, and must spend more time interviewing with a proctor or instructor after completion of the remedial test. Another cost associated with repeated testing involves staff time and energy; the proctors must come to class more often in order to make themselves available to the remediating students, they must spend time in class correcting tests after they are completed, and additional information must be added to the record keeping system. Finally, there are the material costs themselves; more classroom space/time is required to provide room for those remediating, more cost expended in lighting and heating the space utilized, more paper is needed for the tests themselves, more clerical time is used printing them, and the volume of waste paper increases proportionately. It may be in everyone's best interests to (now that the student has the opportunity to remediate) discover techniques which make remediation less likely, without jeopardizing the student's grade. It was the purpose of the present experiment to try to do so, utilizing two different approaches.

The problem may be analyzed in terms of stimulus control. Initially the student's written answers to the study questions are in front of him (after extraction from the book or article), and the student's responses are made in their presence, controlled by the textual stimuli before him. Gradually the student removes the answers from his presence, and his answering response comes to be controlled by the questions alone. On the test then, the student (if the self-programmed study procedure has been effective) correctly emits the appropriate responses in the presence of the questions alone once more, and receives a good mark.

This analysis is certainly not novel; the advocates of programmed

texts (Skinner, 1968) adjust the content of programmed text "frames" such that by the time the text calls for a response on the student's part he is ready to answer it without mistake. Others have examined the fine grain development of stimulus control, concluding that grosser trial and error shaping may in fact be the less preferred method to employ when a fading technique is possible to use (Terrace, 1966; Sidman and Stoddard, 1967; Storm and Robinson, 1973). If the paradigm for the development of stimulus control also applies to the studying process, perhaps a technique which incorporates elements of the fading procedure may likewise facilitate the student's performance on tests.

In the first approach used it was therefore decided to attempt a direct application of fading procedures to the testing process itself. It was reasoned that an "intermediate" or "faded" form of the examination (i.e., a quiz or practice test incorporating elements of the regular examination) might control positive responding (i.e., the student might answer the intermediate test more correctly) better than the final test itself, given that the student had in fact answered the study questions and studied a little previously. An intermediate form of the exam was administered to all those students in the "fading" group, consisting of questions found on the regular test, plus items that were suspected of "helping the student answer" the questions. Such items were worded in the form of "hints", partial answers, alternate phrasing of the question, and often included information as to how many parts a correct answer would contain, problems to "watch out for", and how much detail was required. (For samples of such faded examinations, see Appendix A.) If "errorless" responding could be controlled by this faded test, it was hypothesized that performance

on the regular unit test would be increased, thus possibly decreasing the number of students requiring remedial testing.

In the second approach, a stimulus control procedure based on the concept of modeling was used. During the test correction process in a PSI structured course the proctor typically spends time with the student telling him what was wrong with the answers so marked, advising the student as to what needs to be done instead, and giving him positive feedback for generally good answers. If however, the student has the correct answers written in response to his study questions and simply failed to emit them in the test situation, mere reiteration of these answers by the proctor is not likely to benefit the student any more than if he simply reread and studied again from his written notes. An unpublished literature review by Semb supports this position. If on the other hand the student was studying from an incorrect model in the first place, then the proctor must spend time telling each student individually what the correct answer is. Unless the student has studied from a model that is essentially correct, his imitation of that model is likely to be judged inadequate. Courses taught on PSI lines provide models that are essentially correct, but those models are presented, often, unfortunately, by the proctor correcting the test, when it is too late for the student to do anything about it (except to emit it when next he tests). Therefore the second approach examined the effects of shifting model presentation time from after test taking, to before test taking. That is, students were required to participate in a discussion session after they had completed their study questions but before they took

the first examination for each unit. During this discussion session each study question was answered (correctly) and students were given the opportunity to ask about the various materials they had read. Neither is this approach novel: in fact, Peters (1974) had applied this logic to, and experimentally evaluated the effects of using prequiz "monitoring" (as he calls it) within, two courses taught at C. W. Post college. Positive results were reported, and they were of a magnitude that was very encouraging.

The variables of interest in this study were therefore: the number of tests needed to meet criterion (the dependent variable), the fading program, the discussion program, a typical PSI program, and a combination of the fading and discussion programs (the independent variables). A "practice exam" program was compared with the fading program to assess the relative contribution of practice effects alone to the fading procedure. Comparisons between the various procedures were made to assess their relative effectiveness in reducing the amount of remediation required by the students in the course.

METHOD

Students

Forty-six of the forty-eight students participating in the course activities also participated in the experiment to its completion. The course in which the experiment was carried out was an upper division undergraduate psychology course in Applied Behavior Analysis. Most of the students were psychology majors at Western Michigan University. They registered for the course without prior

knowledge that experimental conditions would exist. Of the two students not participating in the experiment, one did not attend class after the first week, and the other was not taking the course for credit.

Setting and Sessions

The experiment was conducted in two ordinary college classrooms, with a number of tables and chairs but no special apparatus. One room was a tiered lecture room where students took their tests; in the other room tables and chairs were arranged to facilitate the relatively private grading of papers and discussion between the proctor and the student, and, on Mondays, the furniture was arranged to allow seating for the discussion group as well as for the grading of papers. The academic portion of the class (there was a practicum portion as well) was scheduled to meet four days per week, in 50 minute periods running from 4:00 to 4:50 pm.

Procedures

Before the experiment, the various tests and treatments to be used were developed, and proctors were selected and trained. Proctor training consisted of one session during which they played the role of a proctor while the experimenter (an experienced proctor) commented on their approach and suggested effective methods to them. Before each new unit the two graduate assistants, the three proctors, and the instructor previewed the upcoming test and discussed possible problems, arriving at an agreed upon set of answers to be used for grading purposes. Texts for the course were: The Principles and Procedures of Behavior Modification (Sheldon, Sherman, Stokes and Wolf, 1974) and chapters 14 and 15 of Child Behavior: Learning and Development

(Sheppard and Willoughby, 1975). The study questions contained in those materials were not changed, and served as the pool from which tests in the course were developed. These questions were placed at the end of each section of textual material and called for various sorts of answers; most of these answers were of the brief (one to three sentences) essay type, but at least one question per unit called for a longer response, often, to describe a behavioral procedure or program that might be designed to handle the sort of behavior problem germane to that discussed in that particular text section. Each unit of study was comprised, usually, of two or three sections of textual material (e.g., journal articles or essays) and their accompanying study questions. There was a laboratory requirement to be met in this course; students were assigned to a behaviorally based service setting within which they were to perform various clinical functions at an apprentice level. However, no experimental evaluation pertaining to the present experiment was made regarding those laboratory activities.

During the first days of class the students were oriented to the course, asked to submit their overall and psychology grade point averages (GPA), tested over the formal course procedures (not the experimental ones), and given an entrance exam. This entrance exam was composed of questions requiring a general knowledge of behavioral principles; it called for the student to describe hypothetical behavioral procedures necessary to handle behavior problems, and it called upon the student to supply various research designs necessary to answer certain elementary experimental questions. During the last part of the first week the students were assigned to eight experimental

groups, each group composed of a nearly equivalent strata of students randomly assigned to treatments. (The assignment of students to groups will be discussed below in more detail.) Students in the various groups were told to come prepared to take the first unit examination on Monday, Tuesday, or Wednesday, depending on their group assignment; they were also given a schedule describing the activities they were to participate in during the first three units of the course with respect to the different teaching situations, and on what days they were to appear in class to engage in them. The consequences of engaging in these activities, or failing to, were detailed for the students on this schedule, and will be explained below in more detail. Before each new phase of the experiment appropriately revised schedules were given to the students. At this time the students were informed that certain variations in teaching techniques would be employed in the course as part of a continuing effort to improve it. It was explained to them that none of the methods involved more academic or emotional risk than normally found in courses taught at the undergraduate level at Western. They were advised not to permanently adjust their out of class commitments on the basis of their experience in the first few weeks of class, because later in the course they might be required to come to class more or less frequently within the previously scheduled boundaries. No instruction was given as to the exact nature or theoretical underpinnings of the various conditions, but the students were told that if they were curious they could do no harm by speculating, and that they would be debriefed at the end of the course.

Experimental Design

During the second week of class the various teaching techniques

were instituted. The students had previously been assigned to the eight groups on the basis of their rank in terms of scores on the entrance exam, their psychology, and overall cumulative GPAs. The ranking and assignment was accomplished as follows. First the students were ranked on each of the three measures. Those ranked in the top half on all three measures were then randomly assigned to the treatments. Those ranked in the top half on two of the three measures were then randomly assigned to treatments. Next, those ranked in the bottom half on two of the three measures were assigned randomly. Finally, those ranked in the bottom on all three measures were randomly assigned to the eight groups. This procedure allowed the experimenter to make the rather small groups nearly equivalent in terms of variables usually suspected of being relevant while preserving the randomization feature. It also gave the experimenter a blocking variable of possible use later in various statistical analyses. Depending on which group the student had been assigned to the schedule distributed contained modified descriptions of the following experimental conditions.

Baseline. All students participated throughout in the basic "required remediation" technique, with the other teaching situations being superimposed upon this baseline as the design required. As in most PSI styled courses, this meant that the student was required to master the unit (or pass it in this case, at a level of 100% correct on a unit test or remedial) before he could continue to the next unit. As students completed their examinations they gave them to the proctors for grading, and sat down next to them. The proctor commented on the sufficiency of the answers presented, often praising the

examinee for correct answers, and always informing him of how an incorrect answer should be changed if it was to be counted correct the next time. While correcting the test the proctor often presented cues, attempting to evoke a correct answer to test questions which had been both correctly and incorrectly answered. Only written answers were permitted to contribute to the examinee's grade, unlike other versions of PSI (Keller, 1968) because it has been found that students tend to come less prepared to take a test if the proctor is empowered to change the examinee's score on the basis of that student's remarks during the correction process (Hursh, Wildgen, Minkin, Minkin, Sherman and Wolf, 1975). Failure to master the unit the first time the test was taken did not penalize the student permanently, for he could take the test again as many times as was necessary in order to pass it before the next unit was slated to begin (on the following Monday). Special test sessions were held outside of regular class times for those needing them. One other contingency was held constant throughout all conditions. This contingency required that students engage in each imposed activity at intervals of one scheduled class session apart. In other words, students were required to attend class on Monday and engage in their scheduled activity; those students scheduled to participate in any other experimental activity, or those failing to pass a test administered on Monday were required to come in on Tuesday, and so on, until Thursday's activities were completed. Therefore, the primary features of this baseline condition included a PSI type remediation component, provision of study questions, discussion with proctors during test correction, weekly testing, and finally, a pacing component which required the student to pass each unit within one

week's time and to engage in the various instructional activities serially, at one class-day intervals. The student was rewarded for participating by being granted the opportunity to continue in the course. All this was explained to the students during the orientation week, with primary emphasis placed upon the importance of making contributions to the field of their choice (viz., psychology) and the benefits, educationally, of being involved in and learning how to design, an instructional program that runs smoothly.

Fading component. Those students experiencing any faded examination were administered this exam one day before they were to take the regular unit examination. The faded test included those questions to be asked on the unit test (the unit test consisted of a preselected portion of the study questions found at the end of each textual section) and a separate sheet which contained the above mentioned hints and prompts. The experimenter determined what sort of hints and prompts were to be presented by roughly estimating what items would be sufficient to ensure 100% performance by students who had probably written their own answers to the study questions already, but not including so much material that a student who had not been exposed to the texts in the course could successfully complete the test prompted by the material on the sheet of hints. In answering the items on the test then, the students were exposed to a faded version of the answer. However, the students were asked to write out the entire answer (using the hint sheet), and not merely fill in the missing portions, in order to be credited with participation in this activity. Students were told that this was just a practice exam not counting toward passing the unit, but that they were to turn their papers in for correction

just as though it was a regular examination. Although it was hoped that the students would all perform at or near 100%, no mastery contingency was imposed on their performance; they were just asked to "do your best". Proctors scored the student's answers on this test in the same way they corrected regular unit tests, discussing items with the student as described above.

Discussion component. Those students in one of the discussion sessions were told to bring to that session their own completed study questions; the completion (without regard to accuracy) of their study questions was a required activity. The discussion leader (usually the instructor) looked over each student's answers to determine whether at least some answer was presented; if the student had not answered a question, he was dismissed from the session. Students missing the Monday discussion session for whatever reason were met with privately to discuss their answers. After all the students had presented their answers and were admitted to the session, they were asked to answer each question aloud for the group. The leader commented on the correctness of their answers, then gave the students the opportunity to ask questions about the text material, which the discussion leader answered, when applicable, from the answer key. The session terminated when all questions regarding the material had been answered correctly.

Practice test component. Those students participating in one of the practice test situations were administered a practice test one day before they were to take the regular unit test. This practice exam was identical to the unit exam and different from the faded exam in that no answers or hints of any sort were provided the students. The

students were told that this was a practice examination, that it would not count toward their having to master the unit, but that they were to participate in the correction process just as though it was a regular examination. No mastery contingency was imposed on this performance; they too were just asked to "do your best". This group was included to assess the relative contribution of practice effects to any performance demonstrated by the faded examination groups.

Discussion and faded exam component. Members experiencing this component participated in the discussion session on Monday with the other discussion group students, and took a faded examination on Tuesday. They then took the regular unit test on Wednesday. In all other respects members of this group experienced the same contingencies as members participating in the other teaching procedures.

Choice conditions. During each of the final four units of the course students were offered a choice as to which of the various instructional situations they wished to participate in. During the first two choice units (except for the two groups which experienced the simple discussion method) students were allowed to pick either of the two instructional methods they had experienced up to that time (every group experienced both the baseline condition and one other). During the final two units all students were allowed to choose any of the instructional treatments used in the course.

Prequiz questionnaire. From units one through nine each student was asked to rate the degree to which he felt "comfortable" taking the test by filling out a brief questionnaire. (See Appendix B for a copy of this questionnaire.) The students were asked to identify which condition they were experiencing on the form, but not to identify them-

selves. The students were asked to complete this questionnaire to grossly determine student reaction generated by the various methods.

Treatment sequence. The eight groups were paired in twos in order to effect the experimental design. Group A first experienced three units of the baseline condition, followed by three units of the faded examination condition, and finally, four units of baseline conditions were instated. Group B first experienced the faded exam condition (three units), followed by the baseline (three units), finally returning to the faded exam condition (four units). All pairs of groups were similarly counterbalanced and reversed. Group C first entered three weeks of the baseline condition, then the practice test condition for three weeks, followed by four weeks of baseline again. Group D experienced these same conditions, but in reverse order. Group E first underwent three weeks of the baseline condition, then three weeks of the discussion procedure, followed by three weeks of baseline conditions. Then, that is, during the tenth week, both groups E and F experienced baseline conditions. Finally, group E was scheduled to undergo two units of the discussion conditions. Group F experienced these same conditions, but in reverse order (with the exception of the identical procedures during week ten). Group G initially underwent the baseline procedures for three weeks, was next exposed to three weeks of a combination of the discussion and faded examination, followed by a return to four weeks of baseline conditions. Group H experienced these same conditions, but in reverse order. Once more, the final four units of the course (each week corresponded to another unit covered) were completed under the choice conditions described above.

Finally, exceptions were made to the rather strict contingencies should some matter warrant the valid excuse of a student from class. Every attempt was made by the experimenter, course instructor, and staff to maintain a generally friendly and relaxed atmosphere in the course.

Measurement and Reliability

All scores generated by the procedures were kept for possible use later.

Reliability of grading by the proctors was determined by sampling about 20% of the tests graded and then regrading them. To determine intergrader reliability the grading sheets used to record the student's scores were compared. Since the grading sheets provided room for the proctor to mark which individual part of an answer was unsatisfactory, comparisons could be made of this part by part reliability, or test by test (did the student pass, or fail?) reliability. In all cases the general formula for determining percent of interobserver reliability was used, or: $(\text{number of agreements} + (\text{number of agreements} + \text{number of disagreements})) \times 100$.

Of the 877 tests (practice, regular, and remedial) taken by the students completing the course, 169 were randomly sampled for reliability purposes, 19.3% of the whole. Examining student answers for part by part accuracy, intergrader reliability averaged (mean) 95% with a standard deviation of about 9.6 points. From the student's point of view it was most important that graders agree not so much as to the correctness of individual parts of the test, but that they agree as to whether or not the student should have passed the test at all.

Intergrader agreement as to whether or not the student demonstrated mastery of the unit averaged (mean) 80%, with a standard deviation of about 13.6 percentage points. In 91% of the cases in which graders disagreed as to whether or not the student should have passed the unit, the student actually did pass, it was the reliability grader who felt the student should not have passed. Further, of the tests sampled for reliability purposes only 3 actually failed (but the reliability grader thought the student should have passed), and these instances occurred during the first two units only.

RESULTS

Treatment comparisons. The performance (that is, the number of tests required to pass each unit) of each group is contrasted in Fig. 1 with that of the group that experienced the same treatments but in a counterbalanced order. The figure legend for Fig. 1 may be found on page 34, and Fig. 1 may be found on page 35. Treating each compared treatment as a variable and pooling data across comparison groups, the experimenter used a one way analysis of variance (since autocorrelation was not regarded to be a problem when sessions were spaced at weekly intervals) to determine the probability of statistically reliable differences between teaching methods. Examining overall PSI performance for groups A and B (mean = 2.03) and comparing it with the overall faded exam performance of those two groups (mean = 2.25), a difference significant at the 0.05 level was not found ($F =$

1.823, $df = 1, 17$). Similarly, a significant difference was not found between the overall PSI performance of groups C and D (mean = 1.92) and the overall practice exam performance of those two groups (mean = 2.24; $F = 0.716$, $df = 1, 14$). A significant difference was found ($F = 2.482$, $df = 1, 23$) at the 0.01 level when comparing the overall PSI performance of groups E and F (mean = 1.82) and the overall discussion performance of those same groups (mean = 1.26). It was assumed that there was no overall statistically significant difference between the PSI performance of groups G and H (mean = 1.82) and the discussion plus faded performance of those two groups (mean = 2.00), because similarly small differences between the other various groups were likewise insignificant. It is worth noting however, that there was no variability in the discussion plus faded results, that is, whichever group or individual experienced that condition required only two tests to achieve the 100% mastery criterion. It was likewise assumed that no significant differences between the practice exam condition (mean = 2.24) and the faded exam condition (mean = 2.25) would have been found.

Returning to the performance of groups E and F, recall that the overall PSI performance averaged (mean) 1.82 and overall discussion performance averaged (mean) 1.26; this represents a 32% difference. Only during units two and three were the results equivocal, when the groups experiencing PSI conditions performed slightly better (but not significantly better than; using a t test, $p < .75$), and the same as (respectively) the discussion group. Figure 2 illustrates the difference between the discussion and the PSI conditions with respect to

the number of tests taken as a function of each condition, summed across the 11 experimental units during which comparable results were obtained. The figure legend for Fig. 2 may be found on page 36, and Fig. 2 may be found on page 36a. As is shown, the total number of tests taken by those experiencing discussion conditions was 83, as compared with the 120 tests taken by those during PSI conditions. Since there were six students participating in one of these two conditions at any given time, and since 11 comparable experimental units were experienced by these students, \bar{N} was equal to 66 in each case, yielding the average performances described above.

It is worth noting that individual trends support those of their respective groups, that is, group data are fairly representative of the performance of individuals within those groups. In fact, nine out of the 12 students experiencing the discussion versus PSI comparison displayed a difference of 0.05 or more. In other words, when the performance of each student was examined and each student's mean performance during discussion conditions was compared with his mean performance during PSI conditions, it was found that most students took the equivalent of $\frac{1}{2}$ test more, on the average, during each unit he experienced PSI conditions. The other three out of the 12 students did not do worse during discussion conditions than during PSI conditions, but they displayed a difference averaging (mean) only 0.19 between conditions.

Four individual records, two matched pairs of students from groups E and F, illustrating pairs whose performance is most typical and least typical of their respective groups' performance, are shown in Fig. 3. The figure legend for Fig. 3 may be found on page 37, and Fig. 3 may be found on page 37a. As can be seen, students eight and ten were quite representative of the group data presented in Fig. 1, displaying only two instances where discussions yielded results equal to that of PSI conditions. In all other contrasting units, the performance of the discussion student (regardless of which student experienced that treatment) was twice as efficient as that of his PSI counterpart. Students 29 and 35 were not very representative of their respective groups' performances. There were five instances of equal performance per unit, two of those showing PSI to be successful in producing mastery on the first try (that is, the student passed on the first test). Furthermore, during the first phase the discussion student needed to take two tests per unit before passing. Finally, one last individual performance is worth noting. Student 44 consistently scored approximately 2.5 tests taken per unit (this student was not in the discussion pair of groups); this student signed up for the discussion condition when given free choice during the last two units, but mean performance remained unchanged.

Efficiency ratio. Although significant differences were not found when the performances of the other various teaching methods were contrasted, those treatments being contrasted with the matched

PSI components were at a "disadvantage" in that there was a lower limit placed on the number of times they had to test per unit. In other words, each of the groups experiencing the faded, practice, and discussion plus faded conditions could not have scored less than two tests per unit, because they were all required to take a practice test of one sort or another (which counted as one test in the scoring arrangement presently used), guaranteeing a performance of at least two per unit. Illustrated in Fig. 4 is the statistically adjusted performance of each group as it experienced the different treatments across the various units. The figure legend for Fig. 4 may be found on page 38, and Fig. 4 may be found on page 38a. To arrive at the efficiency ratio, the mean number of tests actually taken per unit by each group was divided by the minimum number of tests required for each of their respective treatments. For example, a group with a mean performance of two during one of the phases in which two tests had to be taken as part of the teaching procedure is given an efficiency ratio of $2/2$, or unity. Since the denominator for this ratio for all PSI conditions is unity, the resulting transformation does not change them; their mean scores appear the same. Since the denominator for the other three above mentioned components is two, their transformed scores change into half of what their original performances were. Thus the differences between compared treatments appear to be quite large. However, it can be seen that the differences between the faded, discussion plus faded, and the practice treatments are quite small when examining their respective efficiency ratios.

Choice phases. During the eleventh and twelfth units, when members of all groups (except for the discussion group and its PSI partner comparison group) could choose to enroll in either one of the two treatments they had experienced, every student chose to experience the PSI format. During the thirteenth and fourteenth units when students from all conditions could choose any one of the teaching situations used in the course, three students chose to participate in the discussion session in unit 13, and one student chose to participate in the discussion session during unit 14. The rest chose to experience the PSI format.

DISCUSSION

It appears that use of a prequiz discussion during which all the items to be included on a weekly test are correctly answered and discussed, can significantly decrease the number of remedial tests taken during those units. The decrease is statistically significant, but also of practical import, since the discussion groups needed to take on the average, 32% fewer tests than their matched PSI groups did. However, during the last two weeks of the course when it became possible for any student to choose the instructional program of his preference, only 4% of the students chose to participate in the Monday discussion group, while the remainder preferred the PSI conditions. A systematic correlation between success in these instructional treatments and three preexperimental measures (psychology and overall GPAs, and scores on a course entrance exam) was not found. Likewise, a significant treatment effect was not found, when the treatment

consisted of providing students with a sheet of prompts during a practice test in the hope that it would facilitate errorless performance on both the practice test and the regular unit test which followed the next day (this was the fading treatment). It was seen also that students involved in a discussion plus faded exam program always passed the unit test at the earliest opportunity, that is, on the first regular test, after participating in a discussion session and taking a faded practice test the following day. The reliability index most directly related to the major dependent variable (that is, agreement between graders as to whether the student should have passed the unit test or not) averaged 80%. No systematic bias of results was suspected of being caused by this level of reliability.

Failure of fading. A premise of major import in this experiment was that the faded test would indeed evoke errorless responding, thus increasing the likelihood of errorless responding on the regular test. Hypothetically, such a faded test would "catch" the student "midway" within the student's own fading procedure. That is, it was supposed that the student had already begun a fading procedure of his own as described in the introduction. It was supposed that presentation of the hints and prompts would supplement a partially completed fading process and ensure correct answering on the practice test. In practice, the faded test did not control errorless performance. Picking unit four as being fairly representative of the performance of groups experiencing the fading technique, examination of the raw data revealed that the faded group scored a mean of only 86% on the practice test, not 100% or errorless performance. Since the fading technique did not

control errorless performance, it is possible that: 1) the students in that group did not study until the evening before the regular test (very likely, since the faded exam did not count towards fulfilling the mastery requirement); 2) the sheet of hints itself was not sufficiently well programmed to ensure errorless performance even though students may have studied (from the experimenter's point of view this is not so likely, since he tried to make the hints as obvious as possible after seeing that the students weren't scoring 100% on the faded test); 3) it is possible that in implementing this particular fading technique, the experimenter did not follow the guidelines set by others who have used the technique with success. For example, Terrace (1966) recommends that fading be carried out gradually and in small steps; this experimenter tried to accomplish errorless performance during the manipulation of only two rather abrupt steps, that of the faded test and then the regular test, while supposing that the first one (the step undertaken by the student himself) had already been accomplished. Terrace also recommended that the similarity between stimuli across steps be kept fairly close, but in this experiment there was probably a great deal of difference, physically, between the student's own notes, the faded test, and the regular unit test. In short, it is likely that the groups experiencing the faded treatment program did not perform errorlessly because one or more of the necessary conditions prerequisite to success did not obtain.

Success of discussion. The success of the discussion method in reducing remedial test taking was not an unexpected result. Semb, Hopkins and Hursh (1973) furnished one group of students enrolled in

an introductory course in child development with the correct answers to study questions (from which study questions the tests were partially composed). They found that students performed some four percentile points better on unit quiz questions for which answers had been provided than on quiz questions which had not been so answered. Peters (1974) found that students enrolled in a child psychology and psychological statistics course who had their unit study questions reviewed and discussed prior to taking the unit quiz took only half as many quizzes as the comparison group did, when 100% mastery of each unit was required in order to pass. The present study replicates this effect. It took discussion group members only 1.26 tests on the average to pass a given unit, whereas it took their PSI paired partners 1.82 tests to do so. That is, discussion students took $\frac{1}{2}$ test less than their PSI counterparts, per student each unit. Further, the discussion group answered, on the average, 98% of the test correctly the first time they took it as compared to their PSI partners, who scored 84% correct, on the average, the first time they took the test. It would appear that worthwhile gains are to be obtained by those who use, and partake of, this sort of discussion session. In the present study however, only 4% of the students able to choose to participate in the discussion sessions did so. That not many students chose to participate in the discussion sessions could have been due to at least three factors. First, participation in the session may have been aversive for some reason; perhaps the student didn't enjoy interacting with the discussion leader (this is quite unlikely, since nearly every student in the course rated the instructor as being fair, friendly and knowledgeable). Second, students may have had other

conflicting demands to meet (the choice under scrutiny here was made available only during the final two weeks of the semester), and hence chose to allocate their time to fulfill those needs.

Third, as can be seen in Fig. 1, the tests during the final units were being passed more quickly than previously, even without the use of discussion sessions. Since difficulty of the tests declined to those levels, students would have no need for participation in the discussion sessions. Hypothetically, there are at least four reasons which help account for the success of the discussion treatment in reducing remediation. First, since students were required to fill out their study questions prior to admittance to the discussion session, while their matched PSI partners were not so required, the completion of the study objectives per se might have been the factor functional in decreasing the amount of test taking. However, since test graders were instructed not to answer or correct questions which the student left blank, it is quite likely that those students participating in the PSI section would have had to at least try to find the answers for themselves, instead of relying on the test grader to provide the answers for them. Therefore, the discussion and PSI groups probably both had completed the study questions, although without empirical verification this remains only a strong probability. Second, it may have been that the test graders tended to mark tests as being 100% correct on Tuesday, without regard to the real merits of the test; or, graders may have known "who was in" the discussion groups, and, biased for one reason or another, tended to give those people better grades regardless of their true performance. A large number of tests were

sampled for reliability purposes; intergrader agreement was high, with nearly 95% of the time graders agreeing on the way particular items on the tests were scored, even though agreement as to absolute mastery was lower. So although it was possible that the graders gave credit where credit was not due, with regard to particular test items this was not true. Further, graders could not decrease the amount of time they were required to spend assisting in class simply by reducing their later in the week test correcting load. They could reduce their correcting "load", however, and this may have motivated them to count most every test as being correct on Tuesday. Had this been the case, those in the PSI conditions would have appeared to master the unit on the second test, which they did take on Tuesday; the obtained results are consistent with this hypothesis. Perhaps the only counter-indicating argument is that: 1) the point by point reliability scores were high; and, 2) the results of Semb et al. (1973) and Peters (1974) replicate the findings of this study even though their grading methods were different. It is still possible that the graders "knew" who was in the discussion sections, and graded them differently because of that. Although the experimenter repeatedly told students and graders that he "couldn't tell them how things were turning out", the graders themselves may have suspected the intent of the experiment. The graders could tell which students were in the discussion group, because the discussion group typically met in the room where the graders did their grading, and they showed up (unlike members of the other groups) for their first test only on Tuesday. The factors related to grader bias do therefore have considerable apparent strength in explaining the

results of this experiment. Third, it may be that the discussion method acted in a manner similar to that of the first attempt tests taken by students in the other groups. It may have acted to reduce incompatible emotional responding, served as a means to familiarize the student with the study questions, and given the student just as much practice as other students received in their respective treatments. Fourthly, the success of the discussion participants may be due most specifically to the fact that they were provided with the correct answers to the test, before the test was administered. In effect, this takes the "cheating out of cheating", providing the students in the discussion group with the corrected answers that the other groups received only after taking the test the first time, that is, during the time when the grader sat down with the student and reviewed the test results with him. Perhaps it is this account that most parsimoniously explains the results of this experiment, but again, there remains the possibility that any one of the other above mentioned factors also affected the results.

Use of preexperimental information as correlates. The preexperimental descriptive data gathered and used to assign the students to the various groups in stratified fashion were not used to adjust statistically the performances of the students, "correcting" for preexperimental differences between the students. Although such adjustment would have made for more sensitive analyses, it was clear from the data that no advantage would accrue in this case. First, since most of the treatment groups' results turned out to actually be higher than the PSI groups', adjustment would not have lowered the

scores below those of the PSI groups'. Second, since students "served as their own controls" in a sense, the observed results are meaningful without resort to adjustment. Third, casual examination of the raw data did not turn up any systematic differences as a simple function of level of group strata. However, several individuals did seem to need more tests than their fellow group members did, but these were located slightly below the midpoint, not at one of the extremes of the strata. The evidence suggesting any overall linear relationship between strata position and performance is weak though, since seven of the eight groups had high ranked members who performed nearly identically to their low ranked members (that is, within two tests of each other, summing across the entire experiment). Although preexperimental information gathered may not have aided in the statistical analysis of the data, there was merit in using such available information as a blocking variable in the assignment of students to their respective groups, since with small group sizes sheer random assignment may have resulted in groups which contained nonrepresentative members with respect to these three intuitively important variables.

Prequiz questionnaires. Although the experimenter had planned to distribute and collect the student scored prequiz questionnaires throughout the experiment, this was discontinued in the ninth unit and the results were not used. There was great reason to suspect, from casual observation, that many of the students were marking them without reference to what was being asked; many of the questionnaires were turned in without identification sufficient to sort them into the necessary comparison groups; many of the students in fact

voiced repeated objections to having to fill them out; and the experimenter believed that the choice conditions would allow for a more direct assessment of the variables of interest.

Discussion in an instructional package. That discussion sessions can reduce the amount of remedial test taking by 30% or more is a factor which strongly recommends its adoption as part of existing instructional routines. It is of the utmost simplicity to implement, and probably does not require the actual leadership of the instructor himself to be effective. Where tests are short, an instructor or assistant may review and discuss the students' answers en masse, that is, in a session involving the whole group. When the class is self-paced, the students' answers may be reviewed individually just before test taking. However, when the units are too long to conveniently discuss during one session, or when the material being tested upon is not highly defined, it may be impossible to use this method for decreasing remediation. Also, since one of the alleged goals of education is to teach the student to "answer questions on his own", unless such a goal is explicitly programmed when using the discussion method as described here, the student's rightful task will be considerably reduced, perhaps without ultimate benefit. It should be the task of the instructional technologist not merely to seek methods which reduce variability and increase or decrease performance on specific dependent variables, but to simultaneously locate and define the original sources of variability. In doing so he may discover that even though students "test better", they are unable to arrange their environment to ensure those high performances once the instructor is

no longer available to do so for them. Inasmuch as use of the discussion session makes it unnecessary for the student to arrange part of his own environment so that eventual high test performance will be ensured, to that extent must instructional analysts determine the desirability of his doing so, and that will be the extent to which that skill must now be more explicitly shaped. In short, there will be many instances in which it will be desirable to use methods which decrease "waste"; however, we must be sure that what we call "waste" indeed is such.

Future study. Further study is indeed necessary to determine whether other techniques might reduce the need for remediation, or more generally make education more effective and efficient. Tighter experimental controls should be used in future studies, to eliminate the sources of possible contamination mentioned above, as well. With regard to the variables of interest in the present study, of further interest is the parameter of test difficulty. Perhaps with more difficult tests, the discussion group would have performed no better than other groups. Perhaps the fading group would have performed better if it took the average PSI student three or even four tries to pass the unit test. It does appear, however, that the difficulty involved in developing a convenient fading approach by instructors for use in their own classrooms may be prohibitive. Varying the test difficulty parameter, however, may increase the likelihood of success with such a program if increasing the test difficulty did not change the performance level of the fading group as seen in this experiment; this is not too likely in this author's opinion, however.

--

Of similar interest is the "dose" of discussion needed to achieve the various effects noted; perhaps a very brief and relatively unthorough review would have achieved the same effects. It is of similar interest to discover the effects student prediscussion preparation has on both the number of remedial tests needed to pass a unit and on the effectiveness of the discussion session as a treatment itself. It may be that experimental manipulation of student prediscussion (essentially, prequiz) preparation will achieve all that was accomplished in the present and similar studies, plus gain the added benefit of teaching a student "how to study on his own". Other studies might investigate the use of alternate modes of presenting discussion style treatments. For example, study questions might be provided students which indicate exactly where the answer is to be found in the text. Or, the answers might be distributed in written form instead of orally discussed, as Semb et al. (1973) did.

We have learned that most students who have obtained the correct answers to study questions before the test usually do quite well on the test. In this case, students were given these answers after having made at least a minimum of effort to obtain them on their own. This author believes that now we know that most students do well if they can obtain these answers, we should concentrate on teaching the student to obtain those correct answers on his own.

Figure 1. Mean number of tests corrected for each group, per unit, across treatments.

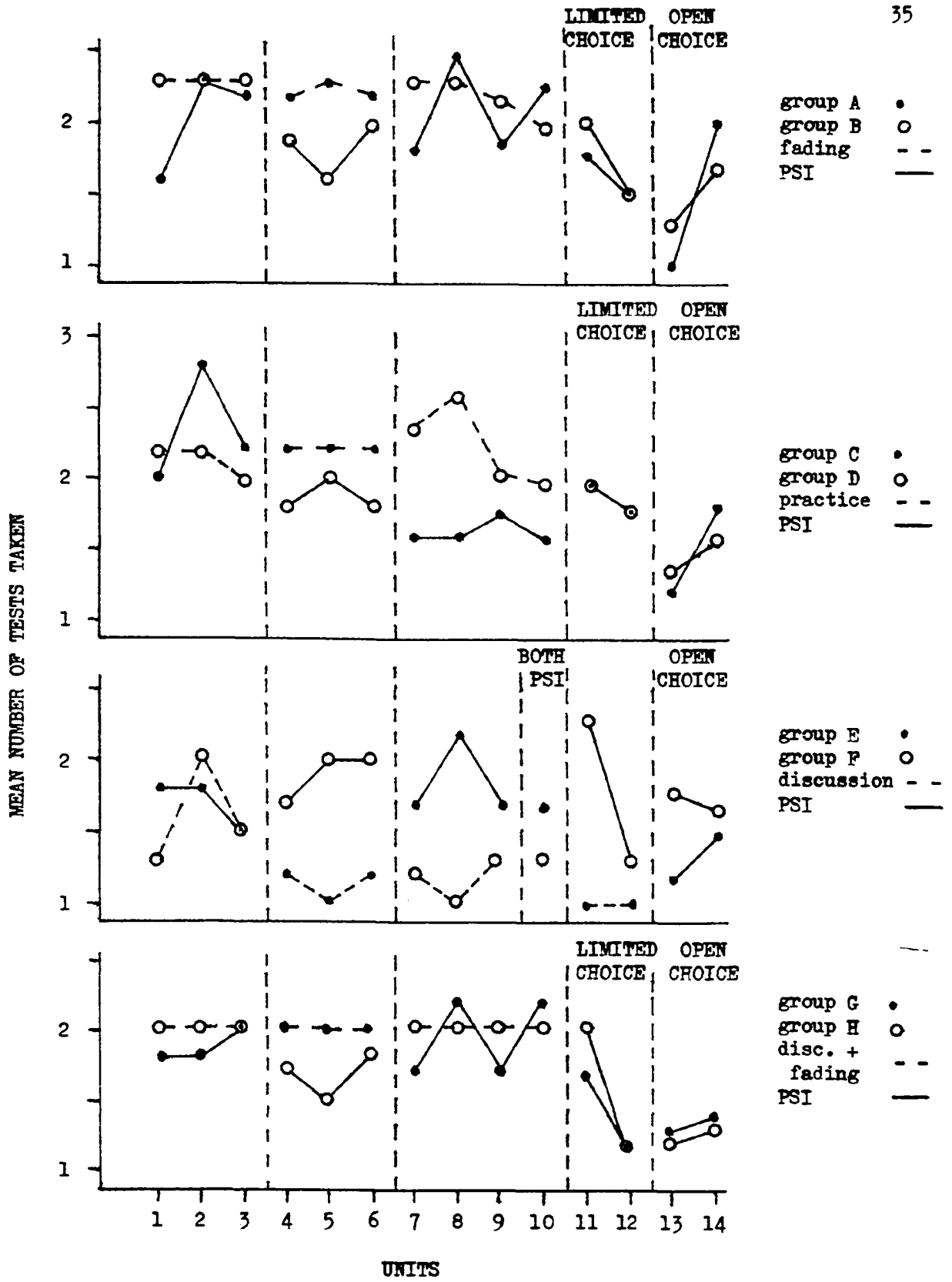


Figure 2. Total number of tests taken by students experiencing the discussion and PSI treatments during the 11 units when those conditions were experimentally comparable.

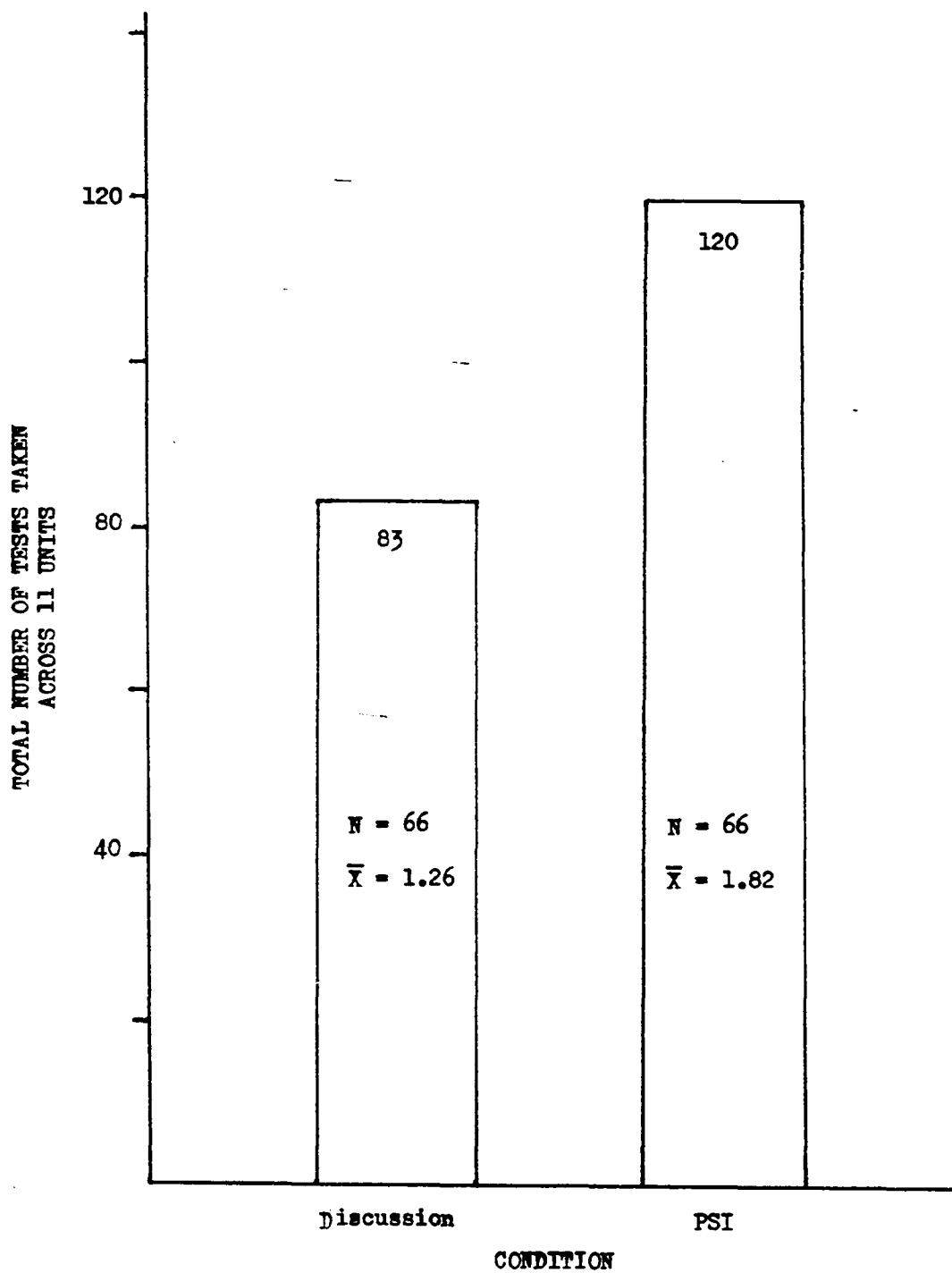


Figure 3. Number of tests corrected for two pairs of students experiencing the discussion vs. PSI treatment comparison. The top graph illustrates the two students least representative of group performance, and the bottom illustrates the performance of students most representative of their groups' performance.

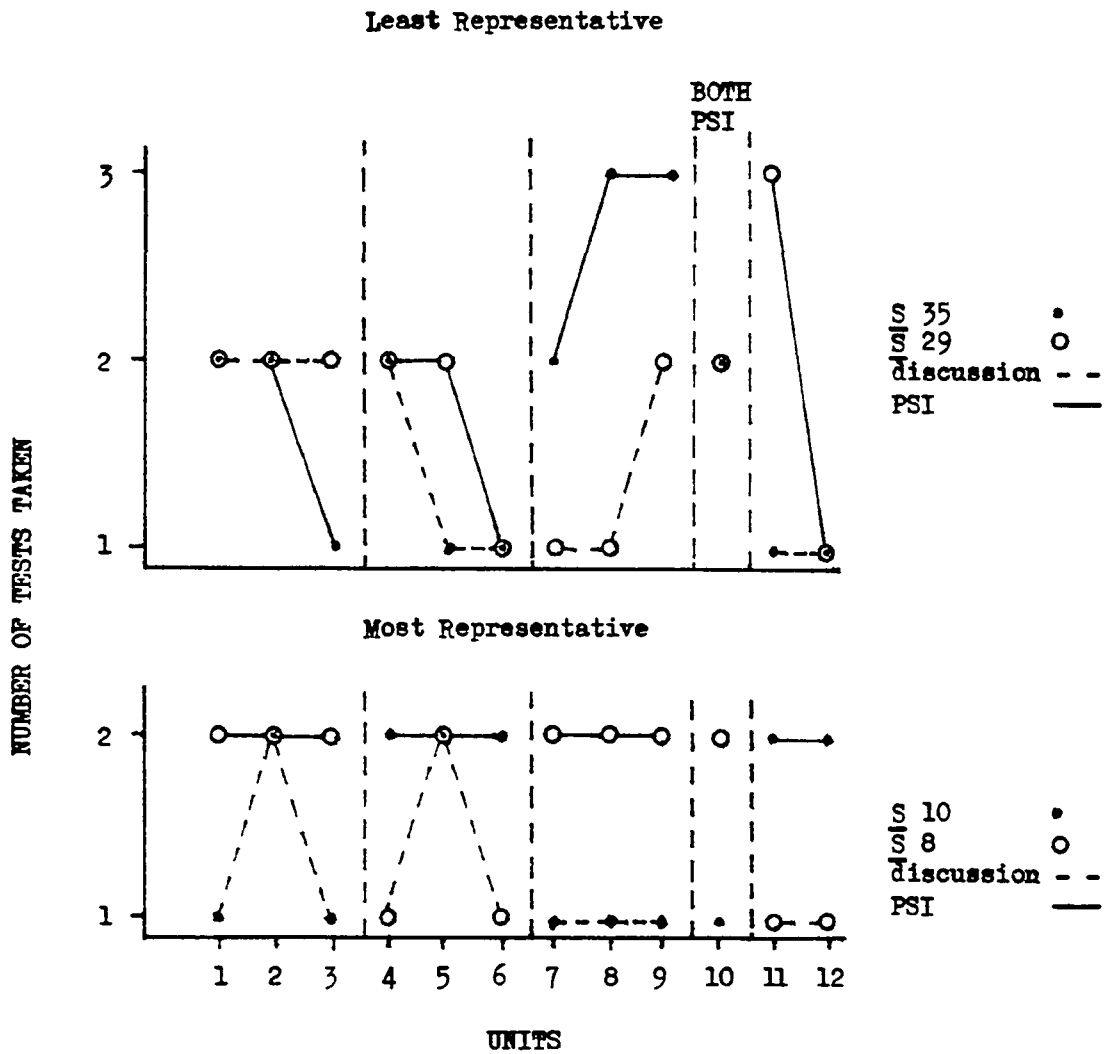
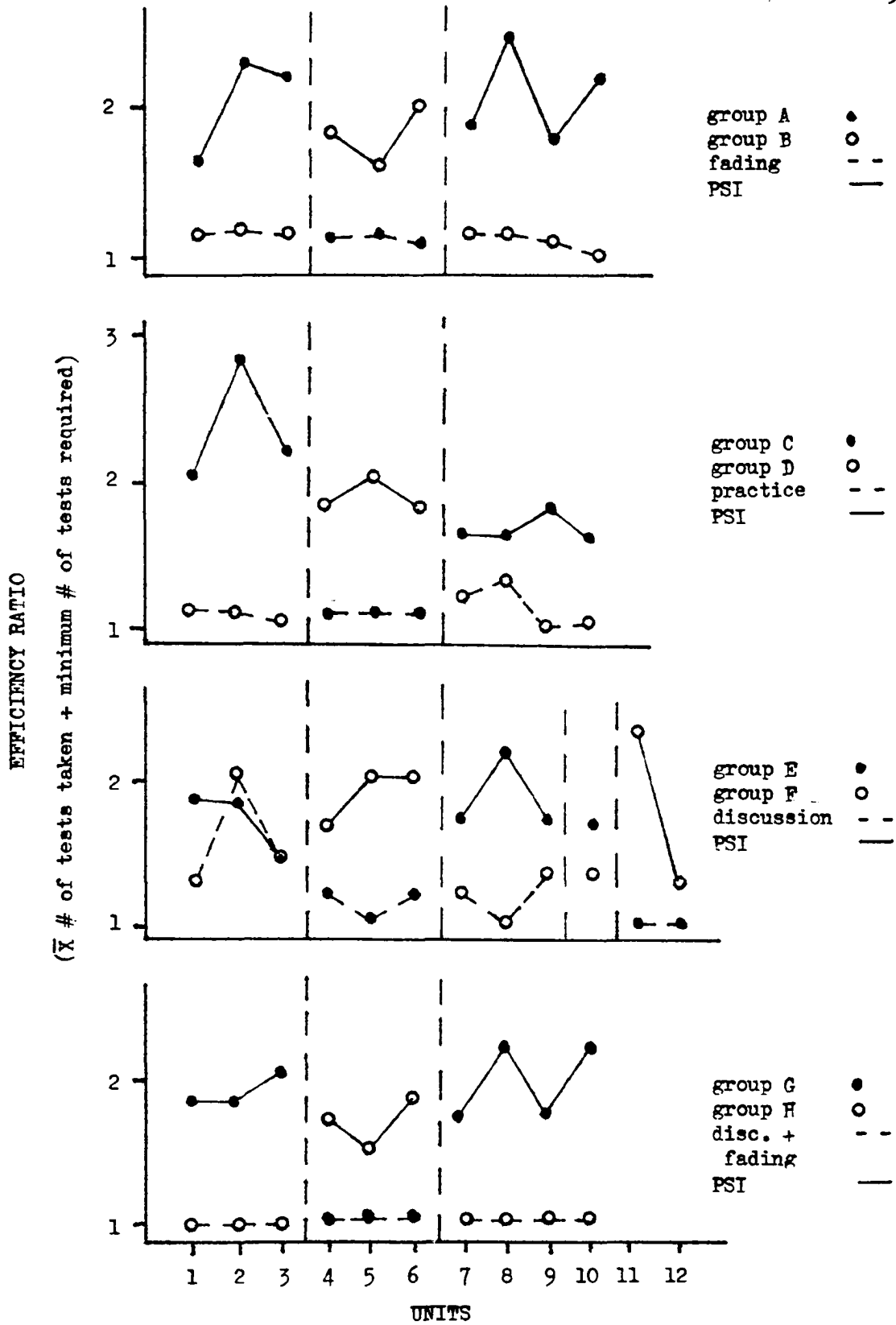


Figure 4. The ratio of the mean number of tests taken by each group, per unit, across units, to the minimum number of tests required to be taken by members of their respective groups.



APPENDIX A
SAMPLE FADED
EXAMINATION QUESTIONS
AND "HINTS"

Sample Faded Examination Questions and "Hints"

Question: One criticism of behavior modification procedures is that they can be used only to teach specific responses to specific stimuli. That is, a result of teaching with behavior modification procedures is someone who will produce only response which have been specifically trained. How do the results with the probes for imitation argue against this?

Hint: Be specific, don't just present the results on probe trials; tell what those results mean in terms of generalization.

Question: According to Risley and Wolf, what are the characteristics of an "ideal" food reinforcer? How might the effectiveness of a food reinforcer be increased?

Hint: They don't want it all over the place, you may need to give it many times, and it has to work. Remember, you can lead a horse to water, and you can increase the chances he'll drink by _____ him of it for 24 hours beforehand.

Question: What criteria did the authors use in choosing responses to be reinforced?

Hint: The responses had to be n _____ and u _____ to the patients.

Question: Why might one be able to argue that giving tokens for the work was providing compensation for the labor?

Hint: Were the tokens worthless?

APPENDIX B

PREQUIZ QUESTIONNAIRE

Prequiz Questionnaire

Please answer these questions to the best of your ability, that is, honestly. Please do not sign your name to this questionnaire; the way you've marked it will remain your own business. Please do try to answer candidly since these questionnaires will be used for important purposes.

1. How worried are you about how you'll do on this test?
not at all 1 2 3 4 5 6 7 very
2. How certain are you that you'll do quite well on this test?
not at all 1 2 3 4 5 6 7 very
3. How well prepared are you for this test?
not at all 1 2 3 4 5 6 7 very
4. How well do you think you'll do on this test?
not at all 1 2 3 4 5 6 7 very
5. How relaxed about the test are you?
not at all 1 2 3 4 5 6 7 very
6. How comfortable are you know, in general?
not at all 1 2 3 4 5 6 7 very

Thank you,

Please feel free to discuss any of these matter with the instructor.

REFERENCES

- Alba, E. and Pennypacker, H. A. A multiple change score comparison of traditional and behavioral college teaching procedures. Journal of Applied Behavior Analysis, 1972, 5, 121-124.
- Corey, J. R., McMichael, J. S. and Tremont, P. T. Long term effects of personalized instruction in an introductory course. Paper presented at Eastern Psychological Association, Spring, 1970.
- Corey, J. R., Valente, R. G. and Shamow, N. K. The retention of material learned in a personalized introductory course. Paper presented at American Psychological Association, Washington, D. C., September, 1971.
- Dalis, G. T. Effect of precise objectives upon student achievement in health education. The Journal of Experimental Education, 1970, 39, #2, 20-23.
- Dustin, D. S. Some effects of exam frequency. The Psychological Record, 1971, 21, 409-414.
- Ferster, C. B. Individualized instruction in a large introductory course. The Psychological Record, 1968, 18, 521-532.
- Gallegos, A. M. A study and comparison of experimental pacing and student pacing of programmed instruction. The Journal of Educational Research, 1968, 61, 339-342.
- Gaynor, J. F. and Wolking, W. D. The effectiveness of currently enrolled student proctors in an undergraduate special education course. Journal of Applied Behavior Analysis, 1974, 7, 263-269.
- Green, B. A. Physics teaching by the Keller plan at M. I. T. American Journal of Physics, 1971, 39, 764-775.
- Gustafson, H. W. and Toole, D. L. Effects of adjunct questions, pre-testing and degree of student supervision on learning from an instructional text. The Journal of Experimental Education, 1970, 39, #1, 53-58.
- Hoberock, L. L. Personalized instruction in mechanical engineering. Journal of Engineering Education, 1971, 61, 506-507.
- Hursh, D. E., Wildgen, J., Minkin, B., Minkin, N., Sherman, J. A. and Wolf, M. M. Proctors' discussions of students' quiz performance with students. In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill.; Charles C. Thomas, 1975.

- Johnston, J. M. and O'Neill, G. The analysis of performance criteria defining course grades as a determinant of college student academic performance. Journal of Applied Behavior Analysis, 1973, 6, 261-268.
- Johnston, J. M. and Pennypacker, H. S. A behavioral approach to college teaching. American Psychologist, 1971, 26, 219-244.
- Keller, F. S. A personal course in psychology. Paper read at American Psychological Association, Philadelphia, August, 1963. Published in R. Ulrich, T. Stachnik and J. Mabry (Eds.), Control of human behavior. Vol. I. Glenview, Ill.: Scott, Foresman and Company, 1966.
- Keller, F. S. "Good-bye, teacher...". Journal of Applied Behavior Analysis, 1968, 1, 79-89.
- Koen, B. V. Self-paced instruction for engineering students. Engineering Education, 1970, 60, 735-736.
- Lloyd, K. E. Contingency management in university courses. Educational Technology, 1971, 11, #4, 18-23.
- McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. Journal of Applied Behavior Analysis, 1969, 2, 79-83.
- Miller, L. K., Weaver, F. H. and Semb, G. A procedure for maintaining student progress in a personalized university course. Journal of Applied Behavior Analysis, 1974, 7, 87-91.
- Moore, J. W., Mahan, J. M. and Ritts, C. A. Continuous progress concept of instruction with university students. Psychological Reports, 1969, 25, 887-892.
- Myers, W. A. Operant learning principles applied to teaching introductory statistics. Journal of Applied Behavior Analysis, 1970, 3, 191-197.
- Peters, R. D. Pre-quiz monitoring of study materials improves performance in two PSI courses. Paper presented at the Second National Conference on Behavior Research and Technology in Higher Education, Atlanta, November, 1974.
- PSI Newsletter, Washington, D. C.: Department of Psychology, Georgetown University, 1971-date.
- Semb, G. The effects of mastery criteria and assignment length on college student test performance. Journal of Applied Behavior Analysis, 1974, 7, 61-69.

- Semb, G., Hopkins, B. and Hursh, D. E. The effects of study questions and grades on student test performance. Journal of Applied Behavior Analysis, 1973, 6, 631-642.
- Sheldon, J., Sherman, J. A., Stokes, T. and Wolf, M. M. (Eds.) The principles and procedures of behavior modification. Lawrence, Kansas; Department of Human Development, 1974.
- Sheppard, W. C. and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. Journal of Applied Behavior Analysis, 1970, 3, 5-11.
- Sheppard, W. C. and Willoughby, R. Child behavior: Learning and development. Chapters 14 and 15. Chicago: Rand McNally Co., 1975.
- Sherman, J. G. A permutation on an innovation. Paper presented at American Psychological Association, Washington, D. C., September, 1971.
- Sidman, M. and Stoddard, L. T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. Journal of the Experimental Analysis of Behavior, 1967, 10, 3-15.
- Skinner, B. F. The Technology of Teaching. New York: Appleton-Century-Crofts, 1968.
- Storm, R. H. and Robinson, P. W. Application of a graded choice procedure to obtain errorless learning in children. Journal of the Experimental Analysis of Behavior, 1973, 20, 405-410.
- Terrace, H. S. Stimulus control. In W. K. Honig (Ed.), Operant behavior: Areas of research and application. New York: Appleton-Century-Crofts, 1966.