An Examination of the Differential Effects of Retesting and a Mastery Criterion on Student Performance in a PSI Course

Carol M. Sundberg

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Psychology Commons

Recommended Citation
Sundberg, Carol M., "An Examination of the Differential Effects of Retesting and a Mastery Criterion on Student Performance in a PSI Course" (1977). Master's Theses. 2291.
https://scholarworks.wmich.edu/masters_theses/2291
AN EXAMINATION OF THE DIFFERENTIAL EFFECTS OF
RETESTING AND A MASTERY CRITERION ON STUDENT PERFORMANCE
IN A PSI COURSE

by

Carol M. Sundberg

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 1977
ACKNOWLEDGEMENTS

The author thanks the following people for their assistance and support for this research: Drs. Jack Michael and Paul Mountjoy for sharing their insight about the research problem, and for their helpful editorial comments; and Dr. Wayne Fuqua for his helpful research suggestions. Especially important in the preparation of this manuscript was the critical and valuable editorial support of both Tim Wysocki and Dr. Richard Malott. The author particularly thanks Dr. Richard Malott for his consistent encouragement and expert advice since the research began. Also, Benjamin Ober deserves special credit for his continual patience, advice, and theoretical analysis pertaining to this research. Finally, the author acknowledges the staff of the Student Centered Education Project for their overwhelming contribution to the academic training of one behavioral research analyst.

Carol Marie Sundberg
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.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
SUNDBERG, Carol Marie, 1954-
AN EXAMINATION OF THE DIFFERENTIAL
EFFECTS OF RETESTING AND A MASTERY
CRITERION ON STUDENT PERFORMANCE IN
A PSI COURSE.
Western Michigan University, M.A.,
1977
Psychology, general

Xerox University Microfilms, Ann Arbor, Michigan 48106
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>16</td>
</tr>
<tr>
<td>III</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>33</td>
</tr>
<tr>
<td>IV</td>
<td>36</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>38</td>
</tr>
</tbody>
</table>
Personalized Systems of Instruction (PSI) are popular alternatives to traditional teaching formats, as increasing numbers of educators support the PSI approach to education. Literature proclaims superior educational gains for students who participate in personalized, rather than conventional lecture-discussion, systems (Born, Gledhill, & Davis, 1972; McMichael & Corey, 1969; Sheppard & MacDermot, 1970). Cole, Martin and Vincent (1975) report both significantly higher exam scores and course retention as a function of personalized methods.

PSI, or the "Keller Plan", spread from an original use in psychology courses (Keller, 1968) to other subject matters. Applications of PSI to entire college curricula (Sherman, 1976) emphasize the broad scale applicability of this teaching technique. Instructors using PSI often modify Keller's basic design. However, the following components are common to almost every personalized system.

1. Short and frequent reading assignments segment the course material into easy, manageable units. Objectives or study guides accompany each unit, highlighting especially important information.

2. Quizzes over each unit of material replace or supplement standard midterm and final exams. The quiz questions are typically multiple choice and short-answer essay.

3. Self-pacing features allow students to study sequential reading assignments, and take a quiz over each assignment, at a self-determined rate of progress. Thus, each student works through
course material at an individualized pace.

4. Mastery criteria guarantee high-level student performance. These criteria are definitions of minimum levels of acceptable performance (typically 85% or above). Progression to new units of material is usually contingent upon meeting the mastery criterion on preceding units.

5. Retesting provides students with, often unrestricted, additional opportunities to reach this mastery criterion.

6. Students receive immediate corrective feedback on their quiz performance, insuring non-interruption of their retesting and course progress. Frequent measures of student performance on quizzes also provide continuous indications of students' course standings.

7. Written materials, often programmed, convey the most critical information. So, books replace lectures as the main source of instruction, though some courses also incorporate lectures as supplements.

8. Student proctors, or teaching apprentices, conduct many aspects of the course, such as grading and bookkeeping. They also serve as tutors, helping students master the course content.

Users of PSI often adopt all the components of this teaching method. Yet, there is a lack of substantial data to demonstrate the effectiveness of each component (Williams, 1976). Since the personalized method of instruction proves superior to more traditional techniques, educators often assume each component enhances student performance.
Future research may prove that each component of PSI contributes to the success of this system. Still, educators need to examine the consequence of eliminating components, perhaps substituting novel ones. Although studies indicate that the components work well together, research also demonstrates that the deletion of some components does not significantly change student performance (Calhoun, 1976; Fernald, Chiseri, Lawson, Scroggs, & Riddell, 1975). Such research is important, since the elimination of components can increase the cost-effectiveness of the system. Alternative remediation procedures, for example, might ultimately save costs and provide benefits superior to those of the present retesting procedures.

In fact, retesting is one of the more costly procedures characteristic of individualized instruction. Retesting necessitates generating multiple quiz forms for each unit of material. Staff resources must also increase to accommodate extra bookkeeping, test administration, and grading. Retesting places additional demands on students, who must reserve variable amounts of time weekly to meet the retest requirements. Still, PSI courses utilize retesting procedures extensively since they contain mastery criteria. This is done because retesting provides students with additional opportunities to reach the mastery criterion.

Advocates of individualized instruction vary widely in their definitions of mastery criteria. Current trends in the application of mastery criteria diverge significantly from Keller's original
concept of 100% correct responding on quizzes. The use of a mastery requirement by Keller (1968) stimulated many educators to experiment with their own versions. McMichael and Corey (1969), and Born et al. (1972), also incorporated in their studies a 100% correct criterion. Students under the Doom's Day Contingency (Malott & Svinicki, 1969) dropped the course or received a grade of F when they failed to reach 100% accuracy on a unit after successive attempts. Sheppard and MacDermot (1970) chose as their criterion joint agreements of mastery, during interviews between two parties. Miller, Weaver, and Semb (1974) required 90% correct responses. Apart from differences in the application of mastery requirements, educators generally agree on the concurrent use of both retesting and mastery criteria.

Johnston and O'Neill (1973) investigated the influence of mastery criteria and grades. They required students to meet a mastery criterion on each unit of material before progressing to the next unit. The students read and answered items orally, and proctors measured rates of correct and incorrect responding. The mastery criterion defined the minimum number of items students could read and answer incorrectly per minute.

The experimenters examined the effect of three levels of the mastery criterion (high, medium, and low), in the first part of their study. Students earned an A for performing at or above the level defined by the mastery criterion. Johnston and O'Neill found that student performance consistently matched the mastery criterion at each level. Whenever the experimenters changed the mastery
criterion, student performance changed to match the new criterion.

In another part of the experiment, all subjects experienced all three levels of the mastery criterion. However, when the experimenters lowered the mastery criterion from high, to medium, to low, they also lowered the grade associated with that criterion. Students earned an A for meeting the high criterion, a B for the medium criterion, and a C for the low criterion. When the medium or low criterion was in effect, however, students could earn a grade higher than the criterion defined. For example, when the mastery criterion defined a grade of C, they earned an A for high performance, a B for medium performance, or a C for low performance. Johnston and O'Neill found that a medium or low mastery criterion defining a grade of B or C, only partially controlled performance, as many students performed above the mastery criterion thus earning a higher grade.

In an extension of Johnston and O'Neill's work, Semb (1974) investigated mastery criteria and assignment length. Students had to pass all course assignments by meeting a mastery criterion on quizzes. Retesting was available, and students took as many tests as necessary to reach the required mastery level. The mastery criterion defined the percent of quiz questions a student had to answer correctly in order to pass a unit. The experimenter varied both the mastery criterion in effect and assignment length. The experimental conditions included; high criterion-short assignments, low criterion-short assignments, and high criterion-long assignments. Semb designated grades of only A or W (withdrawal), so grades
associated with values of the mastery criterion were constant.

Semb found a high criterion superior in generating successful academic performance on first-attempt unit quizzes, first-attempt review quizzes, and post-tests. Results also favored short to long assignments. Long assignments produced lower performance on first-attempt review quizzes and post-tests, even though long assignments included the same high mastery criterion.

Many researchers fail to discriminate between mastery criteria and retesting procedures. For example, Sherman (1976) excluded retesting in his definition of the distinguishing features of PSI. Literature describing personalized systems does not emphasize retesting; usually mastery alone receives recognition (Williams, 1976; Ruskin & Hess, 1976). This may simply imply that educators regard both elements as the same PSI component. Yet, mastery criteria and retesting procedures might be expected to have separate effects on student performance. Investigations of the independent effects of retesting are rare, unfortunately.

Bostow and Blumenfeld (1972) evaluated two retesting procedures. They required students to take initial weekly quizzes, but weekly remedial quizzes were optional. Students worked under a high-incentive retesting system, or a low-incentive retesting system. Students under the high-incentive retesting system earned few points for poor performance on quizzes. For example, a student earned full points for correctly answering at least 80% of the quiz questions, partial points for at least 55%, but no points for below 55%. Points a
student earned on remedial quizzes were added to points the student earned on initial weekly quizzes. In this way, a student improved a poor initial point standing by earning points on a remedial quiz. Students under the low-incentive retesting system earned points in proportion to the percent of quiz questions answered correctly. For example, a student earned 70% of the quiz points for correctly answering 70% of the quiz questions, 50% of the quiz points for correctly answering 50% of the quiz questions, and so on. A student taking both the initial weekly quiz and the remedial quiz received only the highest of the two scores. In this way, students improved a poor initial point standing only when they earned a higher percent of points on a remedial quiz.

Bostow and Blumenfeld found that the high-incentive retesting system increased the number of students taking remedial quizzes. The students lost a great deal of course points for performing poorly on an initial weekly quiz, so they needed to take the remedial quiz to earn points. Students in the low-incentive retesting system frequently did not take the remedial quizzes, because they did not lose a great deal of points for poor initial quiz performance. Also, they only improved their initial point standing when they performed higher on the remedial quiz.

Bostow and Blumenfeld also found that the high-incentive retesting system increased performance on initial weekly quizzes. The mean initial weekly quiz score for the high-incentive group was higher than the mean score for the low-incentive group on eight out
of nine quizzes. Students in the high-incentive group were probably avoiding remedial quizzes. Or, they wanted to be sure they earned points on the first quiz, since poor performance on both the initial quiz or remedial quiz resulted in no points. On the final exam, students in the high-incentive retesting group performed significantly higher than students in the low-incentive group. The experimenters suggested that students in the high-incentive group scored higher because they participated in more remedial exercises. But this result could also be due to their higher initial quiz performance because of this differing point system. Possibly, remedial quizzes did not even contribute to improved performance on the final exam.

Bostow and O'Connor (1973) sought to further clarify the effect of retesting on weekly quiz and final exam performance. One group of students worked under a high-incentive retesting system, where they received few points for poor performance on quizzes. Weekly remedial quizzes were available to students in this group, and points students earned on remedial quizzes were added to points they earned on initial weekly quizzes. A second group of students took each weekly quiz once. Each student in this group earned points in proportion to the percent of quiz questions they answered correctly, and no remedial quizzes were available to students in this group. The experimenters found no differences between initial weekly quiz scores of students who could take remedial quizzes with students who could not take remedial quizzes. And, the experimenters found no consistent increases between initial weekly quiz means and
mean performance on weekly remedial quizzes for students in the retesting group, suggesting that students were not doing additional preparation for the remedial quizzes. But, on the final exam, students who could take remedial quizzes scored significantly higher than students who could not take remedial quizzes. However, the experimental groups took the final exam under different conditions. The high-incentive retesting system was still in effect for the final exam, so students in this group received little credit for poor final exam performance, although they could retake the exam. Students who could not retake quizzes also could not retake the final exam, but these students did earn points for poor final exam performance.

Phillips and Semb (1976) compared contingency-managed lecture courses with PSI courses. In PSI sections, quizzes were student-paced, 100% mastery-before-progression was a requirement, and proctors helped students with the material. Contingency-managed sections had scheduled lectures, quizzes were instructor-paced, and mastery was not a requirement. Both sections incorporated a high-incentive retesting system similar to Bostow and O'Connors. Although remedial quizzes were available to both contingency-managed students and PSI students, only PSI students had to continue taking remedial quizzes until they reached the mastery criterion. The contingency-managed sections, although allowing retesting, restricted retesting to one remedial quiz. The experimenters found no statistically significant difference between final exam scores.

Phillips and Semb noted that since students working under
high-incentive retesting systems cannot earn points for poor performance, these systems increase the likelihood that students take remedial quizzes to earn additional points. Phillips and Semb suggested that high-incentive retesting systems "may be the key to producing performance comparable to that produced by PSI." PSI incorporates a mastery criterion and students cannot earn points until their performance meets the mastery level; so points are withheld for poor performance in PSI, just as with high-incentive systems. Both systems increase the likelihood that students take remedial quizzes.

But, Phillips and Semb's implication that remedial quizzes contribute to improved performance may be incorrect. Retesting may only delay study behavior since students can perform poorly initially, but earn points by improving their performance on a remedial quiz. Since a mastery criterion or a high-incentive system withholds points for poor performance, students might study more to avoid losing points. But, whether they do this studying before an initial quiz or a remedial quiz may make no difference. Retesting may not contribute at all to improved performance, in fact, retesting might even be delaying it. The critical variable leading to improved performance could possibly be simply the fact that points are withheld for poor performance when a mastery criterion or high-incentive system is in effect.

There is evidence that student performance conforms to a mastery criterion when that criterion defines a grade of A. The
grade of A probably controls student performance considerably. Also, mastery criteria, or similar high-incentive systems appear to increase initial performance on quizzes. But, when students do not perform at the mastery level initially, the fact that points are withheld seems to increase, if not guarantee, their participation in retesting.

But it is unclear how retesting contributes to improved final performance, independently of mastery criteria or high-incentive systems. Retests might not provide any additional "learning" experiences, or generate additional study behavior that helps students master the material. So perhaps the only reason why retests might improve performance is because a mastery criterion is still in effect, and points are withheld for poor performance.

The present study examined retesting procedures independently of a mastery criterion. The experiment investigated the effects of retesting as opposed to no retesting on initial quiz performance, final quiz performance, and final course performance as measured by review quizzes.

The experiment also investigated the effects of a mastery criterion as compared to no mastery criterion, on initial quiz performance, final quiz performance, and final course performance.

Finally, this study examined student preference for the experimental procedures. Students chose to work under any of three combinations; (1) retesting with a mastery criterion, (2) retesting with no mastery criterion, and (3) no retesting with no mastery criterion.
METHOD

Subjects and Setting

The experimenter conducted the study in the Student Centered Education Project, a component of the undergraduate psychology curriculum, at Western Michigan University. The program provides beginning psychology majors and minors with an alternative approach to the one-course-per-semester academic sequence, by allowing the students to take two introductory courses in one semester. The project integrates academic instruction with student involvement, so that the students play a major role in the programming and administration of courses.

Students complete the first psychology course in the two-course package during the first half of each semester. The present experiment began halfway into one semester, after the first psychology course was over. The experimenter, however, conducted a study similar to the present experiment during the first course. Of the 43 students who volunteered to participate in the present study, 18 participated in the first experiment.

The experimenter ranked volunteers for the present study on the basis of their review quiz performance in the first course, and assigned them to one of three experimental groups. Equal numbers of student, those who either had or had not participated in the first experiment, were included in each of the three groups. The resultant experimental groups were made up of 14 subjects in Groups 1 and 3,
and 13 subjects in Group 2. Also, review quiz performance, a variable determining the initial rank ordering of subjects, was approximately the same. Mean percent correct on review quizzes was 95, 94, and 95 in Groups 1, 2, and 3, respectively.

Of the initial 43 subjects, one completed only one-third of the course and received an incomplete. Another subject withdrew from the study after finishing three-fifths of the course work. This student indicated that the experimental condition requiring students to retake tests until they reach a mastery criterion interfered with course work by causing students to fall behind assignment deadlines. Neither subjects' data were included in the results.

Course Structure and Personnel

Graduate and undergraduate assistants supervised the course and teaching apprentices conducted the daily class activities. The course format resembled popular personalized systems of instruction. Frequent short reading assignments, with accompanying study objectives, served to divide the course material into manageable units. And corresponding unit quizzes helped guarantee that students studied every assignment. The texts were Elementary Principles of Behavior co-authored by Whaley and Malott (1971), and Tharp and Wetzel's (1969) Behavior Modification in the Natural Environment.

Students read 29 units of material, approximately 15 to 20 pages each, and took quizzes over each unit. Three quiz forms per unit of material were available with each quiz form containing 5 to 10 questions. Quiz questions were a mixture of multiple-choice, fill-
in-the-blank, true/false, and short answer essay. On the average, 60% of the quiz questions corresponded to study objectives, but questions were not simply restatements of prior objectives.

The course consisted of four content sections. The first three sections covered 6 units of material each, with one remaining section 11 units long. A comprehensive review quiz followed the first three content sections. Study objectives for review quizzes were not available to students, and experimental conditions were never in effect for review quizzes. Quiz questions were similar to the type previously described for unit quizzes, however, no questions were identical to former ones. Each review quiz consisted of 10 to 20 novel questions over concepts presented in the content section.

A course schedule of assignments specified the material covered in each unit, the order in which students were to complete units, and a minimum rate of progress for the completion of every unit. The minimum rate of progress signified the slowest pace at which students could work their way through the course, while still completing all course work in one semester. Each unit of material corresponded to a target date, functioning as a reminder to students that the unit should be completed on or before this date. Students paced their own progress through each content section, but received points based on their rate of progress. Students gained points for every unit completed ahead of the course assignment schedule, and lost points for falling behind.
The course also included a weekly lecture and discussion session. Students earned lecture/discussion points by attending, and by taking multiple-choice quizzes over the reading assignments. During these sessions, students interacted on a less formal basis with course personnel.

Grades for the course reflected four measures; quiz points, pacing points, lecture/discussion points, and final exam scores. Students received a course grade of B for earning 90% or more of the total course points, a C for earning between 80% and 89%, and a D for earning between 70% and 79% of the course points. Students who failed to earn at least 70% of the points received an E.

Unit quiz points were the most critical, constituting 66% of the course point total. Review quiz points equaled 16% of the total, while lecture/discussion points made up the remaining 18%. Pacing points directly increased or decreased each student's point total for quizzes and lecture/discussion. Students earned one point toward their total for completing a unit of material prior to the scheduled target date. But they lost two points from their total for completing a unit after the scheduled target date. Students neither gained nor lost points when finishing work on the specified target date for the assignment.

As a result of pacing points, students could lose 12% of the total course points by completing every unit after the target date. This would drop a course grade by one letter. Students could gain a maximum of 6% of the course point total by completing every unit
before the target date. This would raise a course grade by one half letter.

Final exams functioned solely to the students' advantage. A student's course grade could increase by one letter, but not more than one letter, as a consequence of final exam performance. Course grades, however, would not decrease due to poor performance on the final exam.

To raise their course grades, students had to earn at least the same percentage of points on the final exam as they earned on their cumulative point total for quizzes, lectures, and progress through the course. A student who earned 90% of the total cumulative points possible for the course would receive a B. This student could change the grade to an A by earning at least 90% of the total points possible on the final exam. Similarly, a student who received 70% of the total course points, would earn a grade of D. This grade could be raised to a C if the student earned at least 70% of the final exam points, otherwise it would remain a D.

Experimental Design

The experiment conformed to a within group design. The experimental conditions were:

Retesting with a mastery criterion (R-MC). This condition involved a retest procedure, and pretest monitoring of study objectives. A mastery criterion of 90% was in effect, meaning that students had to earn 90% of the points possible for one unit assignment before progressing to another. But students could
retest over a particular unit as many times as necessary to reach the mastery criterion.

Before students received a second test over a unit, they had to write the answers to all study objectives for that unit. Students could use textbooks to complete the objectives, as written answers had to be 100% correct. They submitted their answered objectives to an apprentice who either accepted the answers or returned them for corrections. An accepted copy of the objectives served as a pass to take the second unit quiz.

Retesting with no mastery criterion (R-NC). This condition also involved a retest procedure along with pretest monitoring of objectives. But no mastery criterion was in effect, and retesting was optional. Students could progress to new unit assignments after completing at least one quiz over the previous unit. Students did not have to earn a specific percentage of points on any unit, but they could chose to retake quizzes whenever they felt their performance was unsatisfactory. Again, there was no limit to the number of times students could retake a quiz. The policy of pretest monitoring of study objectives also applied to any quiz these students took for the second time.

No retesting with no mastery criterion (NR-NC). In this condition, students had to take only one quiz over a unit of material before proceeding to another. But unlike the other conditions, it was not possible to retest. And students did not have to earn a specific percentage of points on any unit, as there was no mastery
criterion in effect. Thus, study objectives were never monitored.

Choice. Following exposure to all three experimental procedures, students entered a choice condition. At the beginning of this last phase of the experiment, each student chose an experimental condition under which to complete the remainder of the course.

All groups experienced each of the three experimental procedures, but the sequence of exposure varied between groups (see Table 1). The four phases of the experiment corresponded to the four content sections of the course. At the end of each content section students entered a new phase in the experiment, and changed to another experimental condition.

Proctors were responsible for updating students' quiz records, kept on 8x5" cards. Each card contained a student's quiz scores, and was color coded to inform proctors of the experimental condition a student was under in each content section. The experimenter collected quiz data directly from these cards.
Table I
Sequence of Exposure to Experimental Procedures

<table>
<thead>
<tr>
<th>Phase</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>R-MC</td>
<td>R-NC</td>
<td>NR-NC</td>
<td>Choice</td>
</tr>
<tr>
<td>Group 2</td>
<td>NR-NC</td>
<td>R-MC</td>
<td>R-NC</td>
<td>Choice</td>
</tr>
<tr>
<td>Group 3</td>
<td>R-NC</td>
<td>NR-NC</td>
<td>R-MC</td>
<td>Choice</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

Initial Quiz Performance

Figure 1 shows mean percent correct on initial quizzes for all three groups in the four phases of the experiment. The first three phases show differences between initial quiz performance of groups experiencing different experimental conditions, and the final phase shows differences between groups experiencing the choice condition.

Effects of mastery criterion. The results of the present study indicate that a high mastery criterion generates slightly superior initial quiz performance as opposed to no mastery criterion. In all experimental phases, students in the retesting-mastery criterion condition performed slightly above (a mean group difference of between 2% and 6%) students in the retesting-no mastery criterion and no retesting-no mastery criterion conditions.

Although the latter two conditions excluded a mastery criterion, the initial quiz performance of students during these conditions was fairly high. This result suggests that conditions other than a mastery criterion were controlling high initial quiz performance in this experiment.

An examination of the relationship between unit quiz points and course grades might explain why performance was high in the absence of a mastery criterion. A student's final point total for the course consisted of the points a student earned for unit quizzes, review quizzes, and lecture/discussion. This total directly affected
Figure 1. Mean percent correct on initial quizzes for all three groups in the four phases of the experiment.
the student's course grade. Since unit quiz points made up 66% of the possible course points, students could pass or fail the course based upon unit quiz scores alone. Although students did not have to reach a mastery criterion on each quiz, they still had to demonstrate fairly high overall performance on quizzes or risk earning a lower course grade.

So, in the absence of a mastery criterion, initial quiz performance was probably under the control of performance criteria defining course grades. In fact, regardless of the experimental condition in effect, students had to average at least 85% correct on units to earn a course grade of A. That is, students who earned all the possible course points available for review quizzes and lecture/discussion, would receive only 34% (160 points) of the total course points; they still needed another 56% (263 points) to obtain a final point total of 90%. Since unit quiz points made up 66% (310 points) of the total course points, students would have to earn 85% (263 points) of these unit quiz points to remain within the A range in the course.

The results indicate that the group mean for the initial quiz performance ranged from 82% to 89% correct when a mastery criterion was not in effect. Thus, student performance on units of material did tend to conform to the course criterion defining a grade of A, even when students were not in the mastery condition. These results suggest that the grade of A controls performance considerably.

Perhaps it is unnecessary to have a mastery criterion in PSI courses
of this sort, since simply associating the course grade of A with high performance standards produces performance comparable to that produced by a mastery criterion.

**Effects of retesting with mastery criterion.** In the present study, group means for the initial quiz performance ranged from 88% to 92% correct when a mastery criterion of 90% correct was in effect. This result is consistent with Johnston and O'Neill's (1973) findings. They showed that student performance conforms to a mastery criterion; especially when, as in the present study, the criterion defines a grade of A. Johnston and O'Neill, however, reported only final performance over units of material. They did not clarify the relationship between a mastery criterion and initial quiz performance.

Semb found that a high mastery criterion produced high initial quiz performance, while a low mastery criterion produced low initial performance. This result suggests that a positive relationship exists between a mastery criterion and initial quiz performance. In other words, when a mastery criterion is in effect students attempt to meet the mastery criterion on initial quizzes. The results of the present study also demonstrate this relationship.

**Effects of retesting with no mastery criterion.** The present study also investigated the effect of retesting on initial student performance when no mastery criterion was in effect. One experimental condition excluded a mastery criterion but allowed retesting; students who performed poorly on an initial quiz had the option to retake the quiz, but were not required to do so. The
results indicate that even in the absence of a mastery criterion, students who could retake tests performed fairly high on initial quizzes. In fact, their initial performance tended to conform to the course grade criterion of an A.

These results on the effects of retesting with and without a mastery criterion, suggest that retesting opportunities do not discourage initial high performance; students performed high on initial quizzes in both the retesting-no mastery criterion and retesting-mastery criterion conditions. Two factors possibly account for this result.

First, students perform high initially in order to avoid retaking quizzes, because the additional effort involved makes it somewhat of an inconvenience. The second and perhaps more plausible explanation can be better understood after further examining the consequences of performing below the mastery level on an initial quiz under the retesting-mastery criterion condition.

Since mastery criterion students needed to master a unit by 90% before progressing, initial performance below mastery meant students had to retake a quiz over the same unit. Not only would this inconvenience the student, but time spent retesting would slow a student's course progress. This could negatively affect the student's course grade, as students lost points for falling behind the target assignment deadlines. Semb's course procedures also included target dates for assignments, and students not meeting these deadlines were required to withdraw from the course. So in
Semb's research, as well as the present experiment, it was to the student's advantage to meet the mastery criterion on an initial quiz.

Students in the present experiment under the retesting-no mastery criterion condition, also had to meet assignment deadlines. It was also to their advantage to perform high on an initial quiz, just as it was for students under the retesting-mastery criterion condition. In both conditions, retesting would slow students' course progress and possibly affect their course grades.

So, what might be concluded from these results is that the inconvenience of having to retake a quiz does not, by itself, control the occurrence of initial high student performance. Another, and possibly more critical variable, is assignment deadlines with aversive consequences arranged for failure to meet these deadlines.

**Effects of no retesting with no mastery criterion.** The present study also examined the effect of no retesting on initial student performance when no mastery criterion was in effect. The results do not show any systematic difference in initial quiz performance between students in the no retesting-no mastery criterion and retesting-no mastery criterion condition. In fact, initial performance was about the same in both conditions throughout all phases. The largest difference between these two groups was only 2%, as observed in the third phase of the experiment.

This result suggests that both conditions had similar effects on initial student performance. Poor initial performance would
lead to aversive consequences in both the retesting-no mastery criterion and no retesting-no mastery criterion conditions. Retesting allowed students to improve poor initial performance, but could cause them to fall behind assignment deadlines, and this would negatively affect their course grade. Poor initial performance of students who could not retake tests also would negatively affect their course grade, since they had no opportunity to improve performance.

Effects of choice condition. In the final phase of the experiment, students chose an experimental condition to work under for the remainder of the course. All students in all three groups chose the retesting-no mastery criterion condition. Probably students preferred this condition for two reasons. First, it allowed students to improve poor initial performance by retesting, thus increasing their course point total. And second, it did not require retesting, so students could avoid decreasing their point total if retesting would cause them to fall behind assignment deadlines.

Final Quiz Performance

Figure 2 shows, for each experimental phase, the mean percent of students in each group whose final unit quiz performance was below 90% correct. Figure 3 shows the mean percent correct for all three groups on final unit quizzes. The results show that in the first three experimental phases, the no retesting-no mastery criterion condition produced the highest percent of students whose final performance was below 90% (see Figure 2). Fewer students
Figure 2. Mean percent of students in each group whose final unit quiz performance was below 90% correct.
Figure 3. Mean percent correct on final quizzes for all three groups in the four phases of the experiment.
in the retesting-no mastery criterion condition failed to attain 90% before progressing to new material. The fact that these two conditions produced about the same initial quiz performance, but showed differences in the final quiz performance (see Figure 3), indicates that students did retake quizzes when retesting was optional.

The retesting-mastery criterion condition produced superior final quiz performance in all phases (see Figure 3). One requirement, however, of this condition was that students meet a 90% mastery criterion before progressing to new material. But this requirement was not consistently enforced as represented by the small percent of students in each phase who failed to achieve 90% while in this condition (see Figure 2).

During the choice condition, an average of 40% of the students in each group failed to reach 90% correct on units before progressing (see Figure 2). Since all students chose the retesting-no mastery criterion condition it can be seen that a higher percent of students in each group failed to attain 90% during the choice condition than during the same condition in an earlier phase. One explanation for this result is that the final experimental phase occurred in the last weeks of the semester. Possibly students did not retake quizzes as this slowed their course progress. Not only would they lose points for being behind, but they did not earn credit for course units they failed to complete by the end of the semester. Progress through the course was more crucial in these final weeks.
Review Quiz Performance

Figure 4 shows the median percent correct for each group on review quizzes in the first three phases of the experiment. The mastery criterion and retesting procedures were never in effect for review quizzes. The results show that in each phase, performance of Group 3 exceeded performance of the other groups, while the performance of Group 1 and 2 was approximately equal. These results indicate that there were no consistent effects of any experimental condition on review quiz performance.

One explanation of these results is that the review quizzes were not a sensitive performance measure. Although all experimental conditions generated high initial quiz performance, final quiz performance was highest under the condition including a mastery criterion; it was second highest when a mastery criterion was absent, but retesting available. Still, review quizzes did not reveal any differential effects of experimental conditions, although these conditions previously produced differences in the final quiz performance of students.
Figure 4. Median percent correct for each group on review quizzes in the first three phases of the experiment.
CONCLUSIONS

There are two main conclusions we might draw from the results of this study. First, the level of performance on an initial quiz for each unit was a function of the requirements for earning the grade of A in the present PSI course. During the condition with the mastery criterion, initial student performance conformed to the mastery criterion defining the grade of A. And, during the conditions without the mastery criterion, initial student performance tended to conform to the somewhat lower overall course criterion defining the grade of A. These results suggest that the grade of A controlled performance considerably.

Although retesting was available in two conditions, initial quiz performance of students during these conditions still conformed to the requirements for earning an A. Two factors possibly accounted for this result. First, the inconvenience of having to retake a quiz. And second, the fact that retaking quizzes slowed students' course progress; this could negatively affect their course grade.

The second main conclusion we might draw from this study then is that retesting was not a critical component of the present PSI course. High performance usually occurred initially in the retesting conditions, possibly because retesting involved aversive consequences. Although the final quiz performance of students during retesting conditions was superior to the final performance of students during the no-retesting condition, review quizzes revealed no differential
effects of these final quiz performance differences.

The first conclusion of this study implies that a mastery criterion is an unnecessary component in PSI courses similar to the one described in this research. Simply associating the grade of A with high performance standards produced performance comparable to that produced by a mastery criterion. The second conclusion implies that retesting also is an unnecessary component. Initial student performance conforms to the requirements for earning an A, when retesting is available, but results in aversive consequences.
REFERENCES


Williams, R. L. "Personalized system of instruction: Where it has been and where it needs to go." Unpublished manuscript, University of Kansas, Lawrence, Kansas, 1976.