4-1980

Programmed Review Tests in a PSI Course

Terry E. McSween

Western Michigan University

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

Part of the Educational Psychology Commons

Recommended Citation

https://scholarworks.wmich.edu/dissertations/2627

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
PROGRAMMED REVIEW TESTS IN A PSI COURSE

by

Terry E. McSween

A Dissertation
Submitted to the
Faculty of the Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
April 1980

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGEMENTS

I extend a very special thanks to Dr. R. W. Malott and Michael J. Dillon for their continued support and the system that helped me complete this project. I also wish to thank Dr. Dale Brethower and Dr. Wayne Fuqua for their input and encouragement over the last three years, and Lynn Wilson for the time we sacrificed to this project. I must also thank Tom Welsh and all SCEP assistants that I worked with over the years, especially those who contributed to the projects discussed in the appendices of this manuscript. I also want to acknowledge the contribution of Judy Lasater who took responsibility for the study time data in Experiment II.

Terry E. McSween
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. 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 material submitted.

The following explanation of techniques is provided to help you understand markings or notations 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 through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International
300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# Table of Contents

**Acknowledgments** ............................................................ 11

Chapter

  I. INTRODUCTION .......................................................... 1

  II. EXPERIMENT I: THE EFFECTS OF PROGRAMMED REVIEW TESTS 3
      Method ................................................................. 3
      Results .................................................................... 9

  III. EXPERIMENT II: A SYSTEMATIC REPLICATION ............... 12
      Method ................................................................. 12
      Results and Discussion ........................................... 16

  IV. GENERAL DISCUSSION .................................................. 20

APPENDIX A ................................................................. 24

APPENDIX B ................................................................. 33

APPENDIX C ................................................................. 44

APPENDIX D ................................................................. 51

APPENDIX E ................................................................. 53

APPENDIX F ................................................................. 59

APPENDIX G ................................................................. 61

APPENDIX H ................................................................. 63

BIBLIOGRAPHY ............................................................... 65

REFERENCE NOTES ........................................................ 69
INTRODUCTION

Over three semesters, the Student Centered Education Project (SCEP) developed a set of programmed review tests that help students acquire an average of 90% mastery of course objectives as measured by criterion-referenced final examinations. An earlier evaluation of the SCEP program revealed an average score of 70% on final examinations which compares with typical personalized systems of instruction (PSI) reported in the research literature (Hursh, 1976) and which provided the impetus for increased efforts to improve student performance. (See Appendix A for a complete discussion of how this research developed from a management by objectives system.) To improve performance SCEP initially began using review tests because previous research documented their effectiveness for improving and maintaining high levels of performance (McSween, Note 1).

Two earlier studies demonstrated the effectiveness of review procedures in PSI courses. Davis (1976) demonstrated that students required to complete two review items on each of 19 unit tests scored higher on final exams than students not required to answer review items, unless the final examination counted heavily toward the students' grade. If the examination was weighted heavily, review students did not score significantly higher on final examinations, but did out-perform no-review students on follow-up tests three to four months later. In another study on review procedures Somb (1976) reported that students required to take tests over review units performed better than students not required to take such tests. In both studies review procedures resulted in higher final examination scores.
While these researchers demonstrated the general effectiveness of review procedures, the present project is the first to attempt to identify review items on the basis of student performance and to repeat those items on review tests in an effort to accomplish a specific high level of mastery. Eight separate experimental comparisons were made during pilot work with review tests; all eight differences favored review tests and five of those attained statistical significance. These data suggested that repeating difficult items on more than one review test was more effective than repeating each item on one review test and lead to the first experiment (see Appendix B for more detail on the pilot project).
EXPERIMENT I: THE EFFECTS OF PROGRAMMED REVIEW TESTS

During the next semester review quizzes were programmed on the basis of student performance data. The review quizzes were programmed in that difficult items recurred on several review tests. In this experiment difficult concepts were repeated on from one to four review tests in direct proportion to the number of students missing each concept in previous semesters. While both of these parameters (item difficulty and number of repetitions) were independent variables in earlier pilot work, sequence effects and other factors confounded the results.

Method

Subjects

Twenty out of twenty-four undergraduate psychology students volunteered to participate in this study by signing informed consent forms during the first week of an accelerated introductory class in applied behavior analysis. After they completed the first three daily quizzes, volunteers were ranked on the basis of cumulative quiz performance, then subjects were assigned by pairs to one of two groups, review or no-review, using a random numbers table.

Setting and Course Description

All subjects were students in the Student Centered Education Project (SCEP), a personalized system of instruction for psychology students. SCEP is an accelerated program in which students complete two sequential courses in a single semester, for a total of seven credit hours for the first two psychology courses required for
psychology majors and minors: The first course was six weeks long, while the second lasted the remaining nine weeks. Students received separate grades for each course. To get a course grade of "A", they had to attain 90% or better on the final examination and also on the cumulative quiz point totals. Students had one opportunity to remediate each daily quiz and also the final examination, which they could do by taking an alternate form of the quiz or examination. The criteria for other grades was not specified.

Concept validation. An earlier study (Wilson & McSween, Note 2) resulted in specification of the study objectives targeted by the SCEP program. In that project, SCEP catalogued all study objectives in both courses and surveyed faculty in Western Michigan University's Department of Psychology as to which principles and concepts were important for psychology majors. This survey determined the concepts measured on final examinations and review tests, yielding approximately 170 "validated" objectives covering more than 225 concepts. Daily tests covered these as well as other objectives.

Daily tests. The students completed an average of five assignments per week. Quizzes were available each weekday except Wednesday, when the students attended a required seminar. Each daily quiz was worth ten points and covered one chapter from one of the course texts. The students received a syllabus at the beginning of the semester that specified the last day each quiz was available (target dates for quizzes), and the objectives for each quiz. Students could take quizzes ahead of target dates; and they could
take a remedial quiz, if they wished, up to 24 hours after target
dates, as long as they completed their first attempts by the target
date. They could voluntarily remediate a daily quiz by taking an
alternate form of that quiz and receive the better of the two scores.

The test items restated the study objectives and generally
asked students to define and illustrate some term or concept, list
some important features or rules, or compare and contrast several
concepts. Such items were generally short answer, requiring one or
more sentences. Multiple choice questions typically asked students
to identify some specific feature of a concept required by an essay
item on the same test, or asked them to identify some feature of the
material not assigned in the objectives; less than 5% of the quiz
items were multiple choice.

Review objectives. All students received review objectives
that also served as final examination objectives. Both groups
received these objectives and additional copies were available
throughout the semester. The objectives for review tests actually
referred to the daily test objectives which accompanied each of the
texts. The review objective sheet listed the book, chapter, objective
number, and concept or key words from the objective.

General Procedures

All students in SCEP took review quizzes except for research
subjects assigned to the no-review group. No students in the no-
review group took review quizzes and all students in the review
group took all review quizzes.

Review Quizzes

Students in the review group took review quizzes after approximately
every fifteen concepts (two to five daily quizzes). The review quizzes normally consisted of two components worth a total of 30 points. The first half of each quiz covered the 15 concepts introduced since the last review quiz and the second half covered difficult concepts occurring prior to the last review quiz. Difficult items were defined as those missed on review tests by more than 20% of the students during the previous semester. The exception was the first review quiz which covered 18 new concepts and was worth 20 points. Review test items were the same short answer items used on daily quizzes. No multiple choice items occurred on review tests.

Depending on the level of difficulty, the concept might occur on from one to three additional review tests; concepts occurred on one additional review test if more than 20% of the students had missed them during the previous semester, on two additional review tests if more than 40% of the students had missed them, and on three additional review tests if more than 50% of the students had missed them. The first three columns of Table 1 present the number of review tests for each level of difficulty and the number of objectives at each level.

When these procedures resulted in more than 30 points on a review test, the 40% and 50% difficulty level concepts could not occur on three or four consecutive review tests; the concepts with the fewest repetitions received priority. The length of review tests and the number of repetitions were scheduled to allow the planned repetition, except for the difficult concepts in the last five units. The review test in the second course repeated concepts from the first.
Table 1

The Number of Objectives at Each Level of Difficulty, the Number of Review Tests on Which Those Items Occurred, and Make-up Examinations.

<table>
<thead>
<tr>
<th>Number of review tests</th>
<th>Level of difficulty (% previous semester's students missing concepts)</th>
<th>Number of concepts at each level</th>
<th>1st exam</th>
<th>2nd exam</th>
<th># of objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0 \leq X &lt; 20)</td>
<td>47</td>
<td>27</td>
<td>20</td>
<td>167</td>
</tr>
<tr>
<td>2</td>
<td>(20 \leq X &lt; 40)</td>
<td>77</td>
<td>33</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(40 \leq X &lt; 50)</td>
<td>20</td>
<td>11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(50 \leq X)</td>
<td>23</td>
<td>5</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>167</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a\) The first-course examination actually had two sections, as described in the text.
Dependent Variables

First course examination performance. The examination at the end of the first course was completely comprehensive and provided two separate dependent measures: performance on easy concepts that only occurred once on review tests, and performance on difficult items that occurred on several review tests. By the time of this examination, many of the difficult concepts had not completed all scheduled repetitions. (These repetitions could occur prior to the second examination.) The fourth column in Table 1 presents the number of items at each level of difficulty. The final examination items were unchanged from review tests.

Second course examination performance. The examination at the end of the second course covered concepts from the first course and the first half of the second course. For purposes of this research, no concepts from the remaining five units were on this examination because many of the difficult concepts from those quizzes did not complete the specified number of occurrences. This examination had four sections, one for each level of difficulty. The last column in Table 1 presents the number of items in each section. The two easiest sections consisted of two concepts selected at random from the appropriate difficulty level, in each unit, while the two most difficult sections included all items at those levels of difficulty because there were so few. Again final examination items were unchanged from daily quizzes and review tests.

Highest overall examination performance. Students could take a second final examination in both courses so the highest overall examination score provided a measure that incorporated "remedial" effects.
Daily quiz performance. The students' percent of total possible daily quiz points provided an overall measure of each group's initial acquisition. For this analysis daily quiz points did not include points earned on review quizzes.

Data Collection and Reliability

Teaching apprentices graded all final examinations without knowing whose paper they were grading or whether the examination was from the review or no-review group. A graduate student not familiar with current SCEP students regraded 25% of the 161 final examinations for each group. The observer agreed with the grading of 95% of the items.

Results

All comparisons favored the review group. The review group scored 19% higher on the first-course examination and 17% higher on the second-course examination. The review group received 5% higher on the first-course best examination score and 6% higher on the second-course best examination score. Table 2 presents the data on final examination scores and indicates statistical significance. The differences consistently favored the review group across both levels of difficulty on the first-course examination and across all four levels of difficulty on the second-course examination. While the small group size made achieving statistical significance difficult, the differences were significant on the most difficult sections of both examinations (two-group randomized block analysis of variance). Furthermore, these differences occurred in spite of equivalent levels of acquisition; the two groups performed at equivalent levels on daily quizzes.
Table 2
Percent Correct on the Examination Following Each Course, on Each Section of the Main Examinations, and on Remedial Examinations, for Each Group in Experiment I.

<table>
<thead>
<tr>
<th>Group</th>
<th>Review</th>
<th>No-Review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Range</td>
</tr>
<tr>
<td>Section</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>85</td>
<td>56-97</td>
</tr>
<tr>
<td>1 Repetition</td>
<td>90</td>
<td>54-98</td>
</tr>
<tr>
<td>2-4 Repetitions</td>
<td>93</td>
<td>58-100</td>
</tr>
<tr>
<td>First-Course Overall (including remedials)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>91</td>
<td>72-98</td>
</tr>
<tr>
<td>Second-Course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall*</td>
<td>89</td>
<td>78-100</td>
</tr>
<tr>
<td>1 Repetition</td>
<td>90</td>
<td>61-100</td>
</tr>
<tr>
<td>2 Repetitions</td>
<td>86</td>
<td>59-100</td>
</tr>
<tr>
<td>3 Repetitions</td>
<td>83</td>
<td>66-100</td>
</tr>
<tr>
<td>4 Repetitions**</td>
<td>94</td>
<td>84-100</td>
</tr>
<tr>
<td>Second-Course Overall (including remedials)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>90</td>
<td>78-100</td>
</tr>
</tbody>
</table>

a All repetitions were not completed by the first course examination.

* p __ .01

** p __ .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The data from this experiment support the use of review quizzes, but it needed replication for two reasons. First, the lack of statistical significance cast doubts on the reliability of the differences, so further replication would increase confidence in those effects. Second, collection of other data such as study time and student evaluations would clarify other aspects of the effects of review tests.
EXPERIMENT II: A SYSTEMATIC REPPLICATION

Experiment II was a systematic replication of the effects of programming four repetitions of difficult concepts on review quizzes. In addition, during Experiment II students completed daily self-reports of their study time, completed evaluations at the end of the semester, and took a follow-up test two months after the end of the course.

Method

Subjects, Setting, and General Procedures

Twenty out of thirty-four undergraduate psychology students volunteered to participate by signing informed consent forms. Subject assignment was the same as in Experiment I.

During Experiment II, SCEP used revised daily quizzes; however, the setting, text books, course design, and general procedures were unchanged. The revised daily quizzes combined old units so that review quizzes always followed two daily tests and were now worth 15 points. Daily units now covered between one-half and three chapters, depending on the number of concepts in those chapters. As a result of this change, students never had to take two quizzes a day, unless they had to take a remedial quiz.

Review Quizzes

Review quizzes now followed every two daily quizzes, still covering approximately 15 concepts. One major difference was that all difficult concepts now occurred on three review tests rather than the staggered number of tests used in Experiment I. This change
was made on the basis of the new performance data; students mastered more of the concepts than during the previous experiment, as evidenced by the second and third columns in Table 3 as compared with those columns in Table 1. The first three columns of Table 3 present the number of review tests for each level of difficulty and the number of objectives at each level. Again the appropriate number of repetitions occurred for all difficult concepts except those in the last five units. Again the review test repeated items from daily quizzes.

First-course examination performance. The first examination was the same as in Experiment I. The difficulty levels of many of the concepts changed from the previous study, as evidenced by the third and fourth columns in Table 3 as compared with the third and fourth columns in Table 1.

Second-course examination performance. The second examination consisted of three components: One consisted of easy concepts—those missed by less than 21% of the students and which occurred only once on the review quizzes. These "easy" items were not randomly selected, rather they had the highest average item difficulty from previous semesters, thus minimizing ceiling effects. The second component included 20 randomly selected items from the 28 items missed by between 21% and 35% of the students during the previous semester. The third component included all items on the first 10 review quizzes (the first 5 were from the first course) missed by more than 35% of the students during the previous semester, for a total of 22 items (see the fifth column in Table 5). On both examinations all items came directly from the daily quizzes and review tests.
Table 3

The Number of Objectives at Each Level of Difficulty, the Number of Review Tests on Which Those Items Occurred, and Make-up Examinations.

<table>
<thead>
<tr>
<th>Number of review tests</th>
<th>Level of difficulty (% previous semester's students missing concepts)</th>
<th>Number of concepts at each level</th>
<th>1st\textsuperscript{a} exam</th>
<th>2nd exam</th>
<th>make-up exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0 \leq x \leq 20$</td>
<td>116</td>
<td>44</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>$20 \leq x \leq 35$</td>
<td>29</td>
<td>21</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>$35 &lt; x$</td>
<td>22</td>
<td>16</td>
<td>22</td>
<td>5</td>
</tr>
</tbody>
</table>

\textsuperscript{a} The first-course examination actually only consisted of two sections, as described in the text.
Highest overall examination performance. Same as Experiment I.
The remedial examination consisted of 10 randomly selected concepts from the population of concepts missed by more than 20% of the students from the previous semester.

Study time. Beginning with the third week and continuing throughout the semester students turned in estimates of the time they spent studying. Due to procedural difficulties during the first two weeks, only about 75% of the students turned in study time for each unit, but all students submitted time estimates for the remainder of the semester.

Other measures. A number of other measures provide a more complete analysis of review effects; these included daily quiz performance, student evaluations, student attrition, and a follow-up test. The daily quiz performance measure was the students' percent of total possible daily quiz points. Students also completed subjective evaluations of review tests, which served as a measure of "consumer satisfaction" with review tests. Then, two months after the course, available subjects took a follow-up test. The last column in Table 5 presents the make-up of the follow-up tests. The concepts for the follow-up test were randomly selected from the second five review tests because several subjects became SCEP teaching apprentices at the end of the course, and therefore might have reviewed materials from the first five review tests prior to the follow-up test.

Subjects were contacted about six weeks after the course and were offered five dollars to take a follow-up test. Subjects who volunteered were asked not to study. When they received the follow-up
tests they were told that the person in each group answering the most questions would receive an additional five dollars. Eight students in the review group and five students in the no-review group took the follow-up test.

Data Collection and Reliability

Prior to grading, advanced teaching apprentices made photocopies of 40% of the final examinations selected in a random manner. A graduate student then graded the photocopied examinations for reliability data. The graduate student was unfamiliar with the research project and had no knowledge of the subjects or experimental groups. The graduate student agreed with the grading on 90% of the test items.

The graduate student who did reliability on the final examinations also served as the primary grader on the follow-up tests. The grader did not put feedback on the tests and recorded points on a separate sheet of paper. Another graduate student then graded 30% of the follow-up tests. Graders agreed on 86% of the test items.

Results and Discussion

As in Experiment I, all comparisons favored the review group (see Table 4). On the first-course examination the review group scored 16% higher than the no-review group, while on the second-course examination they scored 27% higher than the no-review group. Both differences were statistically significant. The review group also received higher best overall examination scores for both courses, though these differences were not statistically significant.
Table 4
Percent Correct on the Examination Following Each Course, on Each Section of the Main Examinations, and on Remedial Examinations, for Each Group in Experiment II.

<table>
<thead>
<tr>
<th>Group</th>
<th>Review</th>
<th>No-Review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Range</td>
</tr>
<tr>
<td>Section</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>First-Course</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Repetition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Repetitions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First-Course Overall (including remedials)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second-Course</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Repetition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Repetitions (Mod Diff)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Repetitions (Diff)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second Course Overall (including remedials)</td>
</tr>
</tbody>
</table>

a All repetitions were not completed by the first-course examination.

b These items were selected on the basis of their difficulty.

* p < .01

** p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Furthermore, the differences in performance on components of the examinations were consistent across both easy and difficult items on the first-course examination and across all three levels of difficulty on the second-course examination. The differences were significant on difficult items (those missed by 20% of previous students) on the examinations following both courses. The second-course examination differences on the one-repetition items were also statistically significant. These differences occurred in spite of the fact that the two groups performed at equivalent levels on the daily quizzes. (The source tables for the statistical analyses are in Appendix E and the daily quiz data are in Appendix F.)

The follow-up test data favored the review group though the difference was not statistically significant. The review group averaged 54% while the no-review group averaged 50%, a follow-up difference somewhat lower than reported by other researchers (Davis, 1976). The present data suggest additional investigations of possible follow-up effects.

The students taking review tests reported studying more than the no-review students, though they studied less for the final examination, suggesting that the higher level of performance exhibited by the review group may be a function of the additional study time programmed by the review tests (see Appendix G). The higher final examination performance on the first attempts by the review group is particularly notable since the no-review group reported studying 50% more for the second-course final examination than the review group.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The students in both groups tended to view review tests as a positive feature of the course. Students in general said that SCEP should continue using review quizzes. Furthermore, while most of the review group reported memorizing review objectives, they also reported that they understood most of the objectives by the end of the course (see Appendix H).

Attrition was again higher for the no-review group. One student in the no-review group withdrew from the study in order to take review tests and another withdrew from the course.

During a brief interview after the follow-up tests, students were asked to describe their study habits. Six of the eight review students reported using flash cards for daily quizzes, review tests and final examinations. In the no-review group, one out of four students used flash cards on daily quizzes and final examinations and another reported using flash cards on the final examinations.
The review tests resulted in mastery of 90% of course objectives as measured by final examinations (before remediation) in marked contrast with the average of less than 75% typically reported by PSI research (Hursh, 1976). Figure 1 presents the average final examination performance on our initial evaluation, and the result of review versus no-review comparisons across three semesters, showing that in all six comparisons the students taking review tests received higher scores than their no-review counterparts. The more detailed analysis of the examinations in the present studies found the effects to be consistent across all eleven components of four examinations. This replication of effects lends strength to the notion that the difference is reliable. In addition to these data, a recent review of the literature suggest that review tests consistently produced significant improvements in student performance in all studies not achieving a high degree of mastery (McSween, Note 1), which is particularly surprising given the low frequency with which teachers use such procedures.

While this project and previous research both clearly support the effectiveness of review quizzes, other studies typically fail to achieve average scores of 90% (although Davis, 1976, reported that one of his groups averaged almost 90%). Data from the present experiments suggest that repetition of difficult concepts on review tests can further enhance the effectiveness of review quizzes and can help achieve final examination scores that average better than 90%. The data from the second-course examinations in both courses
Fig. 5. Initial evaluation and average final examination scores for review and no-review groups across the last three semesters.
Fig. 5. Initial evaluation and average final examination scores for review and no-review groups across the last three semesters.
suggest that four repetitions produced higher levels of mastery than fewer repetitions, in spite of the higher difficulty levels of concepts repeated four times. This effect was consistent across both experiments and suggests the importance of systematically programming review tests on the basis of group performance data, at least when more individualized programming is not possible. Further research might compare such programmed review tests with more traditional review tests using a larger group to reliably demonstrate this effect.

Furthermore, data from these experiments also suggest that review tests may help reduce attrition. In both experiments the no-review group experienced a 20% rate of attrition, with a total of three students dropping the course and one withdrawing from the experiment in order to take review tests. Similar results are not reported elsewhere so this finding awaits further replication. Review tests may reduce the number of students who drop out of their coursework by improving performance.

The review tests seem to arrange more study distributed across the course and decrease the amount of "cramming" necessary to achieve 90% on the final examination. Even if the study time reports are somewhat exaggerated, the no-review group apparently spent substantially more time studying for the final examinations than the review group. Review tests may thus be more humane to the extent that they reduce extensive demands on students' study time during the final weeks of the semester when demands from other traditional classes are greatest. On the other hand, instructors must be careful not to place unreasonable
demands on students' total study time throughout the semester and must weigh the value of review units against the value of providing additional course content.

These data suggest several additional areas for future research. (1) The apparent relationship between the use of flash cards and good student performance suggest that flashcards might be a useful intervention for improving the performance of students who are not mastering course material. (2) The present research maximized practice effects for the review group on specific objectives by simple repetition of related test items; future research might investigate enhancing transfer of training (concept formation) by utilizing novel test items rather than simple repetition. (3) The size of the follow-up differences were much smaller in this study than those reported by Davis (1976). Future research might examine possible factors affecting follow-up performance, including such variables as the remedial examination and time to follow-up which were different from Davis's study. (4) Additional research might also consider the cost/benefits of review tests relative to other programming and remediation efforts. The remedial final examinations in the present study, for example, appeared very effective and only the students who got less than 90% had to take them. Review quizzes might be used effectively in a similar, more individualized manner.

In summary, educators can improve student performances through the use of review tests, and they can get students to achieve better than 90% final-examination performance by repeating difficult concepts on such review tests.
APPENDIX A

SCEP's Management by Objectives
The present research grew out of a larger management by objectives scheme developed by Tom Welsh which had three specific goals. The first goal was to teach students to be effective behavior analysts. SCEP's second goal was to manage the SCEP staff effectively; and the third goal was to teach staff additional staff management, systems analysis, and research skills. In this appendix I will describe the objectives and activities SCEP specified with respect to each of these goals.

Goal I

The first goal, to teach students to be behavior analysts, was subdivided into three academic objectives. The first objective was to develop a list of terminal behaviors for SCEP coursework. Lynn Wilson worked on this objective through a Waldo-Sangren scholarship in the Fall of 1977 and Winter of 1978. Under my supervision, Lynn systematically listed and categorized all of SCEP's materials, and evolved three extensive questionnaires containing all major concepts taught in SCEP's academic program. Lynn and I then distributed these questionnaires to a panel that was composed of nine faculty and three graduate instructors selected because they either taught the course from which those items were selected (in the regular curriculum, not SCEP) or because they were on the curriculum committee for the graduate program in Applied Behavior Analysis. The panel members completed the questionnaires, indicating which concepts our students should know by the end of their training. Summarizing the results of that questionnaire left a list of
"validated" concepts that served as a specification of the verbal repertoire we wished to provide students. (For a more detailed report of these activities, see Wilson and McSween, Note 2.)

The second objective was to use the skills specified by the survey to evaluate the effectiveness of the program. Lynn Wilson then constructed final examinations for SCEP I and SCEP II, and students took those examinations at the end of Winter semester of 1978. (A final examination for SCEP II was ready at midsemester as students completed the first course, Psychology 351.) SCEP I averaged %, SCEP II after the first course averaged %, and SCEP II after the second course averaged %.

The third objective was to develop and implement procedures or changes to improve student performance on the final examinations. This objective lead to the review quiz experiments described in Appendix B and the main text. Here I shall describe a number of the other "programming" activities we engaged in during this time period. I think a description of these activities is important so that the reader does not conclude that we adopted a "sledge-hammer" approach with our review procedures; the review procedures were, rather, a part of our programming activities.

After the evaluation in Winter of 1978 I went through our "validated" list of concepts, looked up the original objective in the course material, and wrote a test item over that objective. The quiz items were then typed on 5" X 8" cards. During the Fall of 1978, I recorded the item difficulty (proportion of students who missed the item) for each item on the back of each card. The typist
typed the review tests directly from the cards.

I then evaluated daily assignments used during Winter of 1978 and, where possible, I dropped units covering only a few concepts and added units covering additional concepts. In this manner the number of concepts covered in SCEP I increased by about 15%, or by about 30 concepts.

Using the cards also made our testing much more comprehensive. The examination used at the end of Winter 1978 was only a sample of the total repertoire. The final examination for 151 in the Fall of 1978 was 100% comprehensive; that is, it tested the entire repertoire. The final examination for 161 in the Fall of 1978 was a more representative sample of the repertoire than the examination of the previous semester because it was based on a random sample of all items.

On the basis of data collected during the Fall, we made a number of changes. First I examined all test items and objectives missed by 20% or more of the students, and re-wrote the test item and/or objective if I thought either was the source of student difficulty. This resulted in 83 "supplemental" objectives (these were different from the supplemental objectives in Appendix B Phase 6) in that they were specific changes or clarifications, rather than examples of student answers). Only a small number of test questions or answers were rewritten, probably less than ten percent.

In addition, the staff and I rewrote all daily tests to include all validated test items for the respective units. The same form of test items occurred on both forms of the daily test. Students
had review objectives prior to the daily quiz, and written instructions with the first set of objectives explained that all concepts would be on daily quizzes. The instructions also told students that they should make flash cards on these items to prepare themselves for the daily quizzes so that they could then review these items on review quizzes.

These changes resulted in substantial improvement in student performance. Tables 1 and 3 in the main text present data on the frequency of items missed by different proportions of students on the review quizzes. The Winter students missed substantially fewer items on review quizzes, presumably due to the additional trials on those test items on daily quizzes as well as the other programming activities. Figure 5 in the main text shows the mean scores from the pre-intervention evaluation and review versus no-review groups on final examinations for the last three semesters in SCEP I. This figure supports the claim in the main section of this text; that review quizzes are an important variable functionally related to student performance.

**Goal II Manage Staff Effectively**

SCEP's second goal is to manage staff effectively. This goal is closely related to the first in that the better SCEP's staff management, the better the overall organization of the course, a dimension that probably affects both the performance and attitudes of the students. This goal had two objectives; specifying performance objectives and developing a management system relating to those objectives.
Starting in the Winter of 1977, the first objective under this goal was to develop a list of performance objectives for each level of staff. I generated a first set of objectives for our paid undergraduate assistants during Dr. Malott’s Systems Analysis Course (Psych. 672). Under my supervision, Linda Tipper continued work on this project on a Waldo-Sangren Scholarship. Working with each level of staff she specified performance objectives for Teaching Apprentices (TA’s) and Advanced Teaching Apprentices (ATA’s).

The following semester (Winter, 1978) Linda empirically validated a system that provided ongoing monitoring and point consequences for those behaviors thereby accomplishing our third objective, which was to implement and evaluate a system to ensure the meeting of criteria specified in the performance objectives. The system Linda evolved suggested that TA’s performed most of their tasks fairly consistently, but ATA’s performed only 40% of their tasks. Linda’s system improved ATA performance to better than 90% task completion per week; but we must question the importance of many of those tasks, given the rate of TA behaviors and since ATA duties primarily involved monitoring and feedback to TA’s.

In the Fall of 1978, under the supervision of Rob Restis, Linda followed up her earlier work, monitoring the success of the system developed the semester before. Her data indicated that the TA’s and ATA’s were completing about 90% of their tasks, though she did not attempt to develop more functional ATA behaviors.

In the Winter of 1979, Linda began specification of Course Assistant (CA) and Graduate Assistant (GA) duties, under the
supervision of Bob Kowalski. She had each CA and GA specify a set of recurring tasks, then experimented with public posting as a means of maintaining performance. While performance improved with public posting, the behaviors which the staff specified were often trivial and typically required process measures rather than outcome measures.

During this period SCEP also made progress at the GA levels, positions undergoing substantial change during this period. In the Winter of 1978, Marilyn Rumph specified a checklist of tasks that occur once per semester for the SCEP II GA position. The following year Jim DeShane and Nancy Hinga developed similar checksheets for SCEP I and SCEP's Educational Technology Laboratory. During the Winter semester of 1980 we further refined those checksheets by specifying tasks for the entire semester, for the new GA positions that were redesigned in the Fall of 1979.

Since the specification of our staff management objectives, SCEP has now written task descriptions for all levels of staff including the program director's position. Our activities in this area continue.

Goal III: Teaching the Staff Additional Skills and Knowledge

SCEP's third goal is to teach staff additional skills in behavioral systems analysis, staff management and research. Some staff are undergraduate students enrolled for credit and to justify this credit (and the student's time) we should be teaching them additional skills. This goal also had two objectives, developing a list of skills and developing (and implementing) training procedures that ensure staff acquire the skills.
During the 1977-78 school year Robin Rumph specified objectives and materials for a number of staff training packages. During that winter, Robin developed a programmed package on writing quiz items and objectives. During the same year Robin and I supervised Amy Rogers as she wrote a Waldo-Sangren proposal to formalize our training, which was accepted and funded.

The following year, 1978-79, Amy developed and validated an extensive staff training package under the supervision of Bob Kowalski. Her project further developed and extensively revised the objectives specified by Robin Rumph. Her package included handouts, lecture notes, tests, and data on student performance for each component. She developed packages on total performance systems, career development interviews, Carnegie, flow-charting, and a variety of other topics.

During the Winter of 1979, SCEP had staff that already completed many of Amy's packages, a situation resulting in our offering a Staff Management Lab II for the first time. I took responsibility for this course which consisted of weekly readings, short writing assignments, and tests. Though the course and assignments were not as closely related to SCEP duties as they should have been, the students taking this course rated it very high. The readings and assignments are on file for future use and recycling. Course materials were largely staff management from JOBM, JABA, and business literature, and PSI articles from JABA, Journal of Personalized Instruction, and previous SCEP research reports.

In the Fall of 1979, SCEP tried to integrate the best features
of the above training programs, but the course still proved too
academic and not well related to staff responsibilities. As with
the other objectives, our efforts continue.
APPENDIX B

Initial Pilot Study with Review Tests
After the evaluation of SCEP's academic performance (see Goal I in Appendix A) I began developing review quizzes to "shape" students' performance for final examinations. The original idea was to have completely comprehensive review tests, a procedure that quickly grew impractical and resulted in several modifications. As a result of modifications the experiment was both overly complex and suffered several potential sources of confounding. For these reasons I did not include it in the main text; I include a brief summary and the results of each phase here, however, because the results consistently favored review tests and a number of the effects were statistically significant.

**General Method Summary**

**Subjects and setting.** The subjects were 38 Western Michigan University students enrolled in the Student Centered Education Project (SCEP).

Subjects volunteered to participate in this study by signing informed consent forms which explained the procedures and potential risks. I assigned subjects to one of two groups, a review group or no-review group. Students assigned to the review group took review tests described below. Students in the no-review group were exempt from review tests.

**Course description.** The course followed a modified PSI format, utilizing eight modules constructed from chapters in Malott, Tilema, and Glenn's *Behavior Analysis and Behavior Modification*, Whaley and Malott's *Elementary Principles of Behavior*, and Sulzer-Azaroff and Mayer's *Applying Behavior Analysis Procedures with Children and Youth*.
In order to complete the course students had to complete a quiz over each unit by a "target date" after which the quiz was no longer available. Each daily quiz was worth 10 points and all tests except those for the Malott, Tilema, and Glenn text were short-answer essay.

Phase I

The first portion of this experiment was an attempt to determine the effectiveness of cumulative review tests during the entire first course. The review group took two comprehensive review tests, then both groups took a comprehensive final examination.

All students took a comprehensive examination at the end of the course. The final examination consisted of 80 items worth a total of 110 points. A majority of test items (78) were short answer essay (of those, eight required examples, three were relational items requiring students to compare and contrast multiple concepts, and the remainder were definitions or one-sentence discussion). The remaining two items required simple fill-in-the-blank type responses.

Summary of results. The review group averaged 84% while the no-review group averaged 72%, a result which is statistically significant.

Phase II

The second phase of this experiment was an attempt to determine if the differences found in Experiment I maintained for the remainder of the semester (nine weeks). There were no additional review tests involving the objectives and quiz items used in this follow-up. I used only "easy" items that were missed by less than 35% of the
students, leaving more difficult items for other experiments
designed to further enhance performance on those items. The
dependent variable for this experiment was a portion of the final
examination for Psych. 161, the second SCEP course.

Results. The review group averaged 89% while the no-review
group averaged 73%, a difference that is statistically significant.
The cumulative review procedures result in better maintenance than
no reviews.

Phase III

The third experiment addressed the effect of a single review
test on the "easy" items, defined as the items missed by less than
35% of the students taking those items on the review test. The
independent variable was a review test over material in the fourth
unit of material. Again, I did not include difficult items in this
experiment because they were used in another experiment. The
dependent variable was performance on "easy" items from Unit 4 on
the final examination.

Results. The review group again outperformed the no-review
group. The mean review group score on Unit 4 items was 78% compared
with 67% for the no-review group, but the difference was not signi-
ficant at the .05 level. This suggests that the improvement or
maintenance effects for the easy items only with one repetition is
not as reliable as for material with more difficult objectives or
with more than a single repetition.

Phase IV

The fourth experiment investigated the effect of a review on
a unit composed of both easy as well as difficult items.
The independent variable was one repetition of all validated concepts in Unit 5 on the review test. The dependent variable was performance on the final examination on items from Unit 5.

**Results.** The review group averaged 89% on the Unit 5 material on the final examination, while the no-review group averaged only 76%, a difference that is statistically significant. This suggests that a single repetition of all items on a review test does result in a reliable performance increase, though I must point out that the material in Unit 5 appears to be easier than the material in Unit 4, as seen by comparing the average Unit 4 examination scores (with easy items only) to the average Unit 5 examination scores (with all examination items). Also the Unit 5 material was closer to the final examination temporally.

**Phase V**

In the fifth phase I asked if a review test immediately prior to a final examination improved performance on the final examination. The independent variable was the last review test over material in Unit 6. The dependent variable was performance on the final examination on items from Unit 6. Students did not have the opportunity to look at their corrected test prior to taking the examination.

**Results.** The review group scored an average of 69%, compared with the no-review group's 64%, a non-significant difference. Apparently immediately preceding the final is an inefficient review procedure, though the lack of significance may in part reflect the lack of feedback. A review test early enough to allow feedback might result in a greater measurable difference between the two groups.
Phase VI

In an effort to further enhance student performance and improve our educational materials dealing with difficult objectives, the next phase was an investigation of the use of incorrect student responses as "supplemental" material to Review Objectives. I ranked difficult items from the 151 final on the basis of the proportions of students missing each item. Difficult items were those missed by 36% or more of the students. All students received objectives with examples of incorrect answers on difficult items from the student's 151 examination. The design for this phase was a 2 X 2 randomized black design comprised of a review group vs. no-review group comparison, and a supplemented items vs. non-supplemented items comparison. The review group took a review test comprised of both supplemented and non-supplemented items. Students used these objectives also for the corresponding section of the final examination that was the dependent variable for this comparison.

Results. Table 5 presents the data on student performance for Phase VI. The difference between the review and no-review groups was statistically significant but the differences between the supplemented items and non-supplemented items were not significant (randomized block ANOVA, p ≤ .05). Simply providing instances of incorrect student responses did not improve student performance.

Phase VII

Because of the pre-existing differences in review and no-review group performances in the comparison in Phase VI, Phase VII replicated that comparison using items and item difficulty data taken

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 5
Average Percent Correct on Different Sections of the Final Examination for Phases VI and VII.

<table>
<thead>
<tr>
<th>Group</th>
<th>Supplemented objectives</th>
<th>Non-supplemented objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase VI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td>83</td>
<td>86</td>
</tr>
<tr>
<td>No-review</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>Phase VII</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td>67</td>
<td>69</td>
</tr>
<tr>
<td>No-review</td>
<td>53</td>
<td>48</td>
</tr>
</tbody>
</table>
from the Unit 4 review test. I ranked and matched difficult items on the basis of student performance on the fourth review test. Students received supplemented objectives and a list of non-supplemented objectives. The fifth review test included an item on each of those objectives. The fifth review test and the supplementation of difficult objectives with incorrect test responses were the independent variables in this phase. Performance on the appropriate section of the final examination was the dependent variable.

Results. Table 5 presents the results of these comparisons. Again the differences between groups was significant but the difference between supplemented and non-supplemented items was not. The group that had the review test outperformed the no-review group, but supplementing objectives seemed to have no effect.

Phase VIII

This phase was simply a "generalization" test, a test of "conceptual" mastery. The test was several sections from Miller's (1975) text. Several situations provided instances of thirty concepts from course materials, and students had to identify those concepts. This test was part of the final examination.

Results. The review group averaged 59% while the no-review group averaged 53%, a difference that was not statistically significant. The overall level of "conceptual mastery" was not very high, suggesting the need for supplemental "concept" programming (Miller and Weaver, 1972) or, at least, the need to systematically program those concepts in lab sites.
General Discussion

The use of review tests in the Fall of 1978 improved the overall level of performance to 84%, including the generalization section, an improvement of 12% over the average comprehensive examination score from the previous semester. This improvement is primarily the result of review procedures, though SCEP also recycled its material during that time.

Some researchers may criticize this conclusion, suggesting rather that my results are simply a function of practice effects. Practice effects were part of the independent variables in these studies, though review quizzes clearly affect student study (see Experiment II). Further, the data suggest that the greater the amount of "practice", the greater the mastery. That is, the data on retention of the "easy" items showed significant review effects for multiple review items (Phase II) while the single repetition of "easy" items in Unit IV did not produce significant effects (Phase III). Further, the significance of the results when using all levels of item difficulty (Phase IV) suggest that a single repetition does have an effect.

The ideal, then, is to provide as much repetition as possible to ensure the students' mastery of concepts. This suggests the repetition of concepts to mastery on an individual basis and, barring that, programming repetition on the basis of group performance.

On the basis of this tentative analysis, I designed a review procedure that programmed the frequency of occurrence for each item.
in direct proportion to the level of difficulty, within the constraints of available time. The actual frequency of occurrence for each level of difficulty was a function of available space on review quizzes. The 50-item reviews used in the last half of the Fall were still too long so I decided to schedule 30-item review tests. Each review test was comprised of 15 new concepts and 15 review concepts. Discounting the first review test, there were 13 review tests providing 195 spaces for review items. I then did a frequency count of items in different levels of item difficult. I used these data to arrive at the frequencies indicated in Table 1 in the Method section of Experiment I. I wrote new objectives for each item missed by more than 40% of the students, and rewrote items that appeared vague or unclear. Students received these objectives as "hints" prior to progressing to the daily unit containing that objective.

To further enhance student performance through "practice effects" I added all review quiz items to both forms of the appropriate daily quiz. Thus students who opted to retake a daily quiz had two repetitions of validated test items prior to the review tests or final examination.

After making these changes, the question remained as to whether these changes would result in an average of 90% mastery, and further, whether performance improvements were the result of the review procedures or simply a function of the other changes made in the course. These reasons led to the research project that constitutes the primary body of this report. I should note that all changes
(new objectives, changes in daily quizzes, recycled test items, etc.) all affected both groups in the Winter experiment; that is, the review tests were the only systematically programmed difference between the groups. Further, the changes made for the Winter were a result of practical considerations of constraints within the SCEP coursework and are not recommended as an optimal arrangement, rather these are changes made in striving for 90% mastery. My hope is that others will model the general experimental approach, not our specific parameters.
APPENDIX C

Source Tables for Experiment I
Table 6
Experiment I -- First-Course Examination
Easy Concepts
(Occurred 1 time on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>253.00</td>
<td>6</td>
<td>42.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>171.50</td>
<td>1</td>
<td>171.50</td>
<td>4.04</td>
<td>.091</td>
</tr>
<tr>
<td>TR X BL</td>
<td>255.00</td>
<td>6</td>
<td>42.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>629.50</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7

Experiment I -- First-Course Examination

Difficult Items

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>645.43</td>
<td>6</td>
<td>107.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>171.50</td>
<td>1</td>
<td>171.50</td>
<td>6.35</td>
<td>.045</td>
</tr>
<tr>
<td>TR X BL</td>
<td>162.00</td>
<td>6</td>
<td>27.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>978.93</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8
Experiment I -- Second-Course Examination
Easy Items
(Occurred 1 time on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>125.44</td>
<td>7</td>
<td>17.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14.06</td>
<td>1</td>
<td>14.06</td>
<td>.60</td>
<td>.46</td>
</tr>
<tr>
<td>TR X BL</td>
<td>163.44</td>
<td>7</td>
<td>23.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>302.94</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9

Experiment I -- Second-Course Examination

Intermediate Difficulty Items

(Occurred 2 times on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>292.94</td>
<td>7</td>
<td>41.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>27.56</td>
<td>1</td>
<td>27.56</td>
<td>1.25</td>
<td>.3</td>
</tr>
<tr>
<td>TR X BL</td>
<td>153.94</td>
<td>7</td>
<td>21.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>474.44</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 10

Experiment I -- Second-Course Examination

Medium Difficulty Items

(Occurred 3 times on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>320.94</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14.06</td>
<td>1</td>
<td></td>
<td>.345</td>
<td>.58</td>
</tr>
<tr>
<td>TR X BL</td>
<td>285.44</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>620.44</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11

Experiment I -- Second-Course Examination

Difficult Items

(Occurred 4 times on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>55.94</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>68.06</td>
<td>1</td>
<td>17.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR X BL</td>
<td>74.44</td>
<td>7</td>
<td>68.06</td>
<td>6.40</td>
<td>.039</td>
</tr>
<tr>
<td>Total</td>
<td>194.44</td>
<td>15</td>
<td>10.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

Average Daily Quiz Performance -- Experiment I
Table 12
Percent of Daily Quiz Points

<table>
<thead>
<tr>
<th>Group</th>
<th>First course</th>
<th>Second course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\overline{X}$</td>
<td>Range</td>
</tr>
<tr>
<td>Review</td>
<td>85</td>
<td>58-99</td>
</tr>
<tr>
<td>No-review</td>
<td>88</td>
<td>80-95</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX E

Source Tables for Experiment II
Table 13
Experiment II -- First-Course Examination
Easy Difficulty Items
(l-repetition)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>2012.44</td>
<td>7</td>
<td>287.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>280.56</td>
<td>1</td>
<td>280.56</td>
<td>1.5</td>
<td>.25</td>
</tr>
<tr>
<td>TR X BL</td>
<td>1269.94</td>
<td>7</td>
<td>181.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3562.94</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 14

Experiment II -- 151 Final

Difficult Items

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>921.75</td>
<td>7</td>
<td>131.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>812.25</td>
<td>1</td>
<td>812.25</td>
<td>45.95</td>
<td>.00</td>
</tr>
<tr>
<td>TR X BL</td>
<td>123.75</td>
<td>7</td>
<td>17.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1857.75</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 15

Experiment II -- 161 Final

Easy Items

(Occurred 1 time on review quizzes)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>2383</td>
<td>8</td>
<td>297.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>722</td>
<td>1</td>
<td>722</td>
<td>2.99</td>
<td>.12</td>
</tr>
<tr>
<td>TR X BL</td>
<td>1933</td>
<td>8</td>
<td>241.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5038</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 16

Experiment II -- 161 Final

Moderate Difficulty Items

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>2994.00</td>
<td>7</td>
<td>427.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2304.00</td>
<td>1</td>
<td>2304.00</td>
<td>5.74</td>
<td>.048</td>
</tr>
<tr>
<td>TR X BL</td>
<td>2812.00</td>
<td>15</td>
<td>401.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8110.00</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 17  
Experiment II -- 161 Final  
Difficult Items

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>2873.75</td>
<td>7</td>
<td>410.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>3721.00</td>
<td>1</td>
<td>3721.00</td>
<td>8.55</td>
<td>.02</td>
</tr>
<tr>
<td>TR X BL</td>
<td>3045.00</td>
<td>7</td>
<td>435.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9639.75</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

Average Daily Quiz Performance — Experiment II
Table 18

Percent of Daily Quiz Points

<table>
<thead>
<tr>
<th>Group</th>
<th>First course</th>
<th>Second course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{X}$</td>
<td>$\bar{X}$</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>Range</td>
</tr>
<tr>
<td>Review</td>
<td>91%</td>
<td>93%</td>
</tr>
<tr>
<td></td>
<td>78-93%</td>
<td>81-97%</td>
</tr>
<tr>
<td>No-review</td>
<td>89%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>68-94%</td>
<td>79-96%</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX G

Student Study Time
Table 19
Total Mean Hours of Study per Student in Experiment II

<table>
<thead>
<tr>
<th>Course</th>
<th>Test</th>
<th>Review group</th>
<th>No-review group</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>Daily</td>
<td>17.5</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>Review</td>
<td>5.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Final Exam</td>
<td>6.5</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>Highest Exam</td>
<td>6.8 (2.1)</td>
<td>7.1 (3.2)</td>
</tr>
<tr>
<td>Second</td>
<td>Daily</td>
<td>34.4</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Review</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Final Exam</td>
<td>12.2</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>Highest Exam</td>
<td>12.6 (3.3)</td>
<td>19.1 (3.5)</td>
</tr>
<tr>
<td>Total study time</td>
<td>92</td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Study time data for the first course began in the third week and represents the last six daily quizzes and the last two review tests.

\(b\) Highest exam study time includes study time of those students who took the remedial examination; the number in parentheses indicates the average amount of study for the remedial examination for students who took the remedial.

\(c\) Total study time does not include study for the remedial examination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX H

Student Evaluation Data
### Table 20

Percentage of Students Responding on Evaluation Items

<table>
<thead>
<tr>
<th>Question</th>
<th>Choice</th>
<th>Review</th>
<th>No-review</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Should SCEP use review quizzes?</td>
<td>Yes</td>
<td>94</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>No response</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>2. What percentage of the objectives did you memorize for review quizzes?</td>
<td>0 - 20%</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21 - 40%</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>41 - 60%</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>61 - 80%</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>81 - 100%</td>
<td>68</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>No response</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>3. What percentage of the objectives do you understand now that you are completing the course?</td>
<td>0 - 20%</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21 - 40%</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>41 - 60%</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>61 - 80%</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>81 - 100%</td>
<td>68</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>No response</td>
<td>6</td>
<td>29</td>
</tr>
</tbody>
</table>

*a* Thirty-one students who took review tests completed the evaluation.

*b* Seven students in the no-review group completed the evaluation.
Bibliography


Ferster, C. B. *The role of review material in continuous programming with teaching machines*. Indianapolis, Medical Center, Indiana University, 1960.
Gates, A. I. Recitation as a factor in memorizing. Archives of Psychology, 1917, 6, 489-496.


Jones, H. E. Experimental Studies of college teaching. Archives of Psychology, 1923, 68.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reference Notes
