MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS (7963.3)
STUDY BEHAVIOR AND PERFORMANCE:
effect of practice and test question similarity

John A. Ellis
Wallace H. Wulfeck, II
William E. Montague
William A. King

Reviewed by
John D. Ford, Jr.

Approved by
James J. Regan
Technical Director

Navy Personnel Research and Development Center
San Diego, California 92152
The study was conducted to test the effect of practice or adjunct questions on learning in a real-world training environment. Subjects were students enrolled in a self-study course at the Navy's Interior Communications "A" School. They were assigned to one of three experimental groups or to a control group. Students in the experimental groups received workbooks; those in the control group did not. The workbooks varied as to the amount of questions included that were identical to those
in lesson tests or in the final test. In the first experimental group, all the workbook questions were identical; in the second group, half of the questions were identical; and in the third group, none of the questions were identical. These groups were subsequently referred to as the ALL, HALF, and NONE Groups. At the end of the course, groups were compared on test and subtest scores, time required, and number of tries on lesson tests. In all cases, the performance of Group ALL subjects was superior. Groups HALF and NONE, who had been exposed to some of the test questions, either in the workbook or on the lesson tests, performed no better on the final test than the control group, who had not. Comparisons on subtests showed that practice questions that are not related to test questions can adversely affect both performance and study time.
FOREWORD

This research and development was performed in support of Navy Decision Coordinating Paper, Education and Training Development (NDCP-Z0108-PN), under subproject P30.A, Adaptive Experimental Approach to Instructional Design, and the sponsorship of the Chief of Naval Operations (OP-01). The objective of the subproject is to develop an empirically-based instructional design support system to aid developers in deciding on instructional alternatives based on costs/benefits and specific resource limitations. The purpose of the present effort is to investigate various instructional remedial methods to determine their effects on student learning. It was approved as part of test bed research by the Experimental Training Programs Policy Board established by a Memorandum of Understanding between the Chief of Naval Education and Training and this Center.

Appreciation is expressed to the staff of the Interior Communications "A" School, Service School Command, Naval Training Center, San Diego, California. Without their assistance and cooperation, this study could not have been performed. Ms. Paula Konoske assisted in data collection and in computer programming. Test item development and analysis of subject matter were conducted under Contract N00123-76-C-2045 to Courseware, Inc.

Results of this research and development are intended for use by the Instructional Program Development Centers, the Chief of Naval Technical Training, and the Chief of Naval Education and Training.

DONALD F. PARKER
Commanding Officer
SUMMARY

Problem

Extensive studies have been made of the effect of practice or adjunct questions on learning. Generally, findings have indicated that, under the proper conditions, they can facilitate learning of material similar to that covered by the practice questions and material unrelated or incidental to that covered by the questions. Because of the laboratory controls employed in these studies, however, the experimental environment is quite different from that of a real-world classroom situation.

Purpose

The purpose of this research and development was to evaluate the effect of practice questions in a real-world training environment in which the instructional materials were designed to teach students to remember specific facts and procedures.

Approach

The subjects were male students enrolled in the Circuits II course at the Navy's Interior Communications "A" School from November 1977 through May 1978. This course is designed to teach students the factual and procedural information prerequisite to their job. The instruction, therefore, is memory rather than performance oriented. They were assigned to one of three experimental groups or to a control group. Members of the first experimental group received a workbook in which all questions were identical to those in lesson tests or the final test; the second group, in which half the questions were identical; and the third group, in which none of the questions were identical. The control group received no workbook. Although members of the experimental groups, hereafter referred to as Groups ALL, HALF, and NONE, could use the workbook in any way they chose, they were directed to the book to answer questions after completing a portion of the lesson materials. Thus, they could become aware of the degree of similarity between workbook and test items.

At the end of the study, the four groups were compared on the following variables: performance on the final test and on subtests made up of items within the test; lesson test scores; time required to complete the course; and total number of tries on lesson tests.

Results

1. Group ALL students performed significantly better than the other students on the final test. There was little difference in the performance of the other three groups, even though those in Group HALF had more exposure to the final test questions.

2. Control group students, who had no workbook, scored significantly higher on items seen once before (in lesson tests) than Group HALF and NONE students, and Group ALL students scored higher than Group HALF students on items seen twice before (in workbooks and lesson tests). There was no significant difference in performance of Group HALF, Group NONE, and control group students on items not seen prior to the final test. These items are comparable to "incidental" questions in the typical adjunct question study.
3. Group ALL students performed significantly better than the others on the lesson tests, followed by those in Group HALF. There was little difference in the performance of Group NONE and control group members.

4. Group ALL students required the least time to complete the course, followed by those in Group HALF, Group NONE, and the control group. They also had the fewest number of tries on lesson tests, followed by those in Group HALF, the control group, and Group NONE.

Conclusions

In a real-world learning environment, practice questions can have a detrimental effect on learning factual and procedural information incidental to the information covered by the practice questions. Further, if they are not related to the test questions, both test performance and study time are adversely affected.

Recommendations

1. If the intent of the instruction is to have students remember important pieces of factual and procedural information (as in the present study), a practice item and a test item should be individual for each piece of information. There should be no practice on information that will not be tested. If time and resource constraints do not permit all important information to be tested, the course material should be prioritized so that the most crucial information can be practiced and tested. The same prescription applies to classroom instruction where the teacher provides practice by oral quizzing or in workbooks, and to other situations in which information is to be memorized. If the intent of the educational program is to transfer learning to new situations, however, practice items would necessarily not be the same as test items.

2. Future research on practice questions should be accomplished in an ongoing training program, and should include an investigation of the effects of different instructional contents, question types, and question frequency.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Problem</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>3</td>
</tr>
<tr>
<td>Subjects</td>
<td>3</td>
</tr>
<tr>
<td>Procedure</td>
<td>3</td>
</tr>
<tr>
<td>Performance Variables</td>
<td>3</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>5</td>
</tr>
<tr>
<td>Final Test Performance</td>
<td>5</td>
</tr>
<tr>
<td>Lesson Test Scores</td>
<td>5</td>
</tr>
<tr>
<td>Study Time Required</td>
<td>8</td>
</tr>
<tr>
<td>Number of Attempts on Lesson Tests</td>
<td>8</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>9</td>
</tr>
<tr>
<td>RECOMMENDATIONS</td>
<td>11</td>
</tr>
<tr>
<td>DISTRIBUTION LIST</td>
<td>13</td>
</tr>
</tbody>
</table>
INTRODUCTION

Problem

The Navy Instructional Program Development Centers, under the Chief of Naval Education and Training, are tasked with analyzing, designing, and developing a large portion of the Navy's technical training courses. The development effort includes writing practice or self-study questions to accompany the instructional materials. Educational psychologists have studied the effects of practice or adjunct questions extensively¹ and have concluded that, under the proper conditions, they can facilitate learning. Because of the laboratory controls employed in these studies, however, the experimental environment is quite different from that in a real-world classroom situation.

Purpose

The purpose of this research and development was to evaluate the effects of practice questions in a real-world training environment. The findings are intended to be used by the Instructional Program Development Centers in formulating guidelines for writing practice questions.

Background

A considerable number of studies have investigated the effects of practice or adjunct questions on learning.² Typically, in these studies, the performance of one or more experimental groups is compared to that of a control group. All subjects in all groups are given a passage of instructional material. Those in the experimental groups are also given a number of practice or adjunct questions in which the variables can be manipulated across groups in a number of ways. For example, the questions may be asked before or after subjects have read the passage, they may be factual or inferential, they may or may not be similar or identical to those included in the final test, or they may vary as to the number included or the frequency of administration. Experimental subjects are usually required to answer the questions without referring back to the material. Finally, all groups are given a final test, and the results are compared to determine whether providing adjunct questions affected performance.

Results of such studies have shown that providing practice questions after—as opposed to before—subjects have read the pertinent section of the reading passage can be effective in some situations. Generally, experimental groups who are given practice questions after they have completed the passage perform better than a no-question control group on final test questions that are similar or identical to the practice questions, as well as on those that are unrelated or incidental to the practice questions. This second finding is important because it shows that providing practice questions after sections of instructional materials helps the student to learn information that is unrelated to that covered in the questions. Presumably, this effect occurs because the practice questions focus the student's attention on the type of question (e.g., factual vs. inferential) and/or the type of information (e.g., general vs. detailed) that will be in the final test.

The implications of these findings for instructional design and development are not clear, primarily because the experimental environment differs greatly from the real-world individualized instructional environment, where students rarely are prohibited from reviewing the instructional materials prior to answering practice questions. In a study by Hiller, students were permitted to check their answers to practice questions. They were required to answer the practice questions before looking back, however, and were not allowed to erase any incorrect answers. This situation produced inferior performance in one condition and no difference in another—relative to a control group—on questions unrelated to the inserted questions. Given these results, it may be that practice questions actually inhibit learning in an individualized learning environment. The present research examines this possibility by varying the number of practice questions that are identical to the test questions in an ongoing technical training course.

---

3See footnote 2.

METHOD

Subjects

The subjects were 289 male students who were enrolled in the Circuits II course at the Navy's Interior Communications "A" School, San Diego, CA from November 1977 through May 1978. The Circuits II course is a 1-week, self-study unit comprised of seven lessons. After a student studies the material for a lesson, he takes a lesson test, which is comprised of both multiple-choice and matching items and is administered on a computer. The passing grade for lesson tests is 70 percent. If the student does not reach this criterion, he must retake the test on the computer. If he fails after three tries, he must take a written test given by the instructor. Test items are not changed for retakes and the student is not given the correct answer. When a student has successfully completed all seven lessons, he takes a final test on the computer. This test includes 50 items, some of which are identical to those included in the lesson tests. The student must answer 64 percent of these items correctly to complete the course.

Procedure

Subjects were assigned to one of three experimental groups or to a control group. All of the subjects proceeded through the course as described above. Those in the experimental groups, however, received a workbook containing practice questions at the beginning of the course. Those in the control group did not receive a workbook.

Although all of the workbooks contained the same number and type of questions, they differed as to the number of questions included that were identical to those included in the lesson tests or in the final test. In the workbook provided to the first experimental group, all the questions were identical; in that provided to the second group, half were identical; and in that provided to the third group, none were identical. Students in the three experimental groups, which will be referred to hereafter as Groups ALL, HALF, or NONE, could use the workbook in any way they chose. The instruction was designed, however, so that they were directed to the workbook to answer questions after completing a portion of the lesson materials. Because students were tested after each lesson, they had an opportunity to become aware of the degree of similarity that existed between practice and test questions as they progressed through the lesson materials and lesson tests. This procedure is somewhat different than the procedure used in the majority of adjunct questioning studies, in which students are only given one lesson to study and one criterion test. In this situation, the student does not have the opportunity to learn the relationship between the practice and test questions.

Although all students in a given class (i.e., those entering each week (about 20)) were assigned to the same group, group conditions were randomly assigned to classes. This resulted in 64 students being assigned to Group ALL, 80 to Group HALF, 84 to Group NONE, and 61 to the control group.

Performance Variables

The four groups were compared on the following variables:

1. Final test performance, including final test score and performance on the following subgroups of items:

   a. Those that had been seen once before (on the lesson tests) by Group HALF and NONE and by control group subjects. (Group ALL subjects had seen them twice before--once on the lesson test and once in the workbook.)
b. Those that had been seen twice before (in the workbook and on the lesson tests) by Groups ALL and HALF.

c. Those that had not been seen before by Group HALF, Group NONE, and control group subjects. This is the measure of performance on questions incidental to the inserted questions.

2. Scores for each lesson test in the order taken.
3. Time required to complete the course (in minutes).
4. Total number of tries on all lesson tests (maximum of three tries per test).
RESULTS AND DISCUSSION

Final Test Performance

As shown in Table 1, Group ALL subjects, who had received workbooks in which all questions were identical to those in lesson tests or in the final test, performed significantly better on the final test than did other subjects (F (3,285) = 57.16, p < .001). There was no significant difference in the performance of HALF, NONE, and control group members, even though Group HALF subjects had been exposed to more of the final test questions than had those in the other two groups.

Final test data were further analyzed by dividing test items into subtests, depending on how many times the various groups had seen the item prior to the final test. The mean performance of the various study groups on these subtests is included in Table 1. Results of an analysis of variance performed on subject results are described below:

1. Subtest comprised of items seen once before--Control group members, who had no workbook, scored significantly higher on this subtest than did members of Groups HALF and NONE (F (2,212) = 5.33, p < .005). A possible explanation for this result is that students in the HALF and NONE groups learned that questions seen before were not likely to be repeated. Therefore, they may not have tried to remember them or they may not have studied the material related to them while preparing for the final test.

2. Subtest comprised of items seen twice before--Group ALL subjects performed significantly better on this subtest than did Group HALF subjects, even though both groups had an equal amount of exposure to the items. This result is consistent with that for the subtest comprised of items seen once before and may be explained in the same way.

3. Subtest comprised of items only on the final test--These items are comparable to the "incidental" questions in the typical adjunct question study. There was no significant difference in the performance of members of Group HALF, Group NONE, and the control group on this subtest. Thus, the presence of practice questions did not facilitate the performance of Groups HALF and NONE relative to that of the control group.

Lesson Test Scores

Figure 1 compares the mean performance of the study groups on the seven lesson tests. It should be noted that subjects were not required to study the lessons in any particular order. Thus, one subject in a particular group may have taken Lesson 2 first; and another, Lesson 7. Since this measure compares performance on the lesson tests in the order they were taken, data from different tests can contribute to a given data point. This did not present a problem in interpreting results, however, since all of the lessons were independent in topic and there were no systematic differences between groups in the order in which lessons were completed.

As shown in Figure 1, the performance of Group ALL subjects was superior to that of the other three. Group HALF subjects generally performed between those in Group NONE and the control group. Both main effects and the interaction are significant: F groups (3,285) = 68.67, p < .001; F lessons (6,1710) = 28.16, p < .001; F interaction (18,1710) = 2.62, p < .001.

It is interesting to note the data for the first lesson. Here the groups order as follows: ALL, HALF, control, and NONE, with large differences between the HALF and
Table 1
Comparison of Mean Performance by Study Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group ALL</th>
<th>Group HALF</th>
<th>Group NONE</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Final test--percent correct</td>
<td>92.22</td>
<td>78.96</td>
<td>76.72</td>
<td>78.25</td>
</tr>
<tr>
<td>2. Subtest of final test questions that had been seen once before (N = 29) (i.e., on the lesson tests) by Group HALF, Group NONE, and the control group--percent correct</td>
<td>N/A</td>
<td>76.00</td>
<td>77.00</td>
<td>81.00</td>
</tr>
<tr>
<td>3. Subtest of final test questions that had been seen twice before (N = 8) (i.e., in the workbook and on lesson tests) by members of Groups ALL and HALF--percent correct</td>
<td>93.50</td>
<td>89.00</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4. Subtest of final test questions seen only on final test (N = 12) by members of Groups HALF, NONE, and the control group--percent correct</td>
<td>N/A</td>
<td>78.00</td>
<td>74.80</td>
<td>74.87</td>
</tr>
<tr>
<td>5. Time required to complete unit (minutes)</td>
<td>364.00</td>
<td>400.23</td>
<td>443.89</td>
<td>495.71</td>
</tr>
<tr>
<td>6. Total number of tries on all lesson tests (minimum = 7, maximum = 21)</td>
<td>7.12</td>
<td>7.74</td>
<td>8.96</td>
<td>8.30</td>
</tr>
</tbody>
</table>
Figure 1. Mean performance of study groups on lesson tests in the order taken.
ALL groups and the control and NONE groups. This result can be explained by assuming that the experimental groups began the course using the workbook questions to guide their study behavior. For those in Group ALL, this was an optimal strategy, since their workbook questions were identical to the test questions. For those in Group HALF, only half of the workbook questions were the same, so they did not perform as well as those in Group ALL. Those in the control group had no workbook and no guidelines; consequently, they apparently tried to study everything and performed almost as well as those in Group HALF. Finally, those in Group NONE found that their guidelines were completely inaccurate, and, not surprisingly, they had the lowest performance. After the first lesson, Group NONE subjects apparently realized that their workbook was misleading and they began to study everything. As a result, their performance became more like that of the control group. It is apparent from this analysis, and from the final test data, that practice questions play a major role in shaping study behaviors.

The general decline in performance for all groups, particularly on the last test taken, can be attributed to two factors. First, subjects tended to leave the more difficult tests until last. Second, the fact that they were permitted to leave class early on the final day if they finished the final lesson and passed the final lesson test produced a "goal gradient" effect.

Study Time Required

Based on the assumption discussed above (i.e., that students begin the course by using their workbook and adjust their study behavior depending on how much it helps them), it can be predicted that (1) Group NONE and control group subjects would require approximately equal time to complete the course, since they would need to study all of the material; (2) Group HALF subjects would require less time because of their exposure to more questions identical to those in lesson and/or final tests; and (3) Group ALL subjects would require the least amount of time, because they had been exposed to all of the test questions. As shown in Table 1, group means, expressed in minutes, were generally as predicted. An ANOVA performed on study time data showed a significant difference among groups: $F(3,285) = 10.18, p < .001$.

Number of Attempts on Lesson Tests

As indicated previously, a student could take each lesson test up to three times on the computer if he continually failed. Thus, the number of tries on lesson tests could range from 7 to 21. Given the previous findings, the groups would be expected to order as follows on this variable: Group ALL, Group HALF, control group, and Group NONE. This assumption was supported, as shown by the group means in Table 1. An ANOVA performed on lesson test data showed a significant difference among groups: $F(3,285) = 26.60, p < .001$. 
CONCLUSIONS

Two conclusions can be drawn from the results from the final test and the subtests. First, adjunct questions in this situation did not help students learn information incidental to the adjunct questions. Second, students in the HALF and NONE Groups apparently realized that the questions in their workbook were not likely to be included in tests; thus, they did not pay as much attention to the lesson test questions as those in the ALL Group, who learned that questions in their workbook would be included in the tests, and the control group, who had no workbook for guidance. As a result, students in the HALF and NONE Groups did not perform as well as the others on questions identical to the lesson test questions. Thus, adjunct questions had an inhibiting effect on their performance.
RECOMMENDATIONS

It is clear that practice questions have important effects on performance, study behavior, and study time. These results, which confirm and extend Hiller's (1974) findings, have direct implications for the design and development of individualized instructional materials. If the intent of the instruction is to have students remember important pieces of information (as in the present study), a practice item and a test item should be individual for each piece of information. There should be no practice on information that will not be tested. If time and resource constraints do not permit all important information to be tested, the course material should be prioritized so that the most crucial information can be practiced and tested. The same prescription applies to classroom instruction where the teacher provides practice by oral quizzing or in workbooks, and to other situations in which information is to be memorized. If the intent of the educational program is to transfer learning to new situations, however, practice items would necessarily not be the same as test items.

Future research on practice questions should be accomplished in an ongoing training program, and should investigate the effects of different instruction content, different question types, and different question frequency.
DISTRIBUTION LIST

Chief of Naval Operations (OP-102) (2), (OP-11), (OP-964D), (OP-987H)
Chief of Naval Research (Code 450) (4), (Code 458) (2)
Chief of Information (OI-2252)
Director of Navy Laboratories
Commandant of the Marine Corps (MPI-20)
Chief of Naval Education and Training (O0A), (N-2), (N-5), (N-9)
Chief of Naval Technical Training (Code 016)
Chief of Naval Education and Training Support
Commander Training Command, U.S. Pacific Fleet
Commander Training Command, U.S. Atlantic Fleet (Code N3A)
Commander, Naval Military Personnel Command (NMPC-013C)
Strategic System Project Office (SP-15)
Commanding Officer, Fleet Combat Training Center, Pacific
Commanding Officer, Fleet Combat Training Center, Pacific (Code 00E)
Commanding Officer, Fleet Training Center, San Diego
Commanding Officer, Fleet Anti-Submarine Warfare Training Center, Pacific
Commanding Officer, Naval Education and Training Program Development Center (Technical Library) (2)
Commanding Officer, Naval Development and Training Center (Code 0120)
Commanding Officer, Naval Aerospace Medical Institute (Library Code 12) (2)
Commanding Officer, Naval Technical Training Center (Code 01E)
Commanding Officer, Naval Damage Control Training Center
Commanding Officer, Naval Education and Training Support Center, Pacific (Code N1B)
Commanding Officer, Naval Health Sciences Education and Training Command (Code 2) (2)
Commanding Officer, National Naval Dental Center (Library)
Commanding Officer, Naval Training Equipment Center (Technical Library)
Commanding Officer, Service School Command, San Diego
Officer in Charge, Naval Instructional Program Development Detachment, Great Lakes
Officer in Charge, Naval Education and Training Information Systems Activity, Memphis Detachment
Officer in Charge, Central Test Site for Personnel and Training Evaluation Program
Director, Training Analysis and Evaluation Group (TAEG)
Director, Career Information and Counseling School, Service School Command, San Diego (Code 3722)
President, Naval War College
Provost, Naval Postgraduate School
Master Chief Petty Officer of the Force, Naval Education and Training Command (Code 003)
Technical Library, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Flying Training Division, Air Force Human Resources Laboratory, Williams Air Force Base
CNET Liaison Office, Air Force Human Resources Laboratory, Williams Air Force Base
Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base
Army Research Institute for the Behavioral and Social Sciences (Reference Service)
Army Research Institute for the Behavioral and Social Sciences (Field Unit--USAREUR) (Library)
Military Assistant for Training and Personnel Technology, Office of the Under Secretary of Defense for Research and Engineering
Commandant, Industrial College of the Armed Forces
Director, Defense Activity for Non-Traditional Educational Support