Navy Personnel Research and Development Center
San Diego, California 92152-6800 TR-92-18 August 1992

Long Term Memory for Different Types of Classroom Knowledge

George B. Semb
John A. Ellis
John Araujo

NPRDC-TR-92-18 August 1992

Long Term Memory for Different Types of Classroom Knowledge

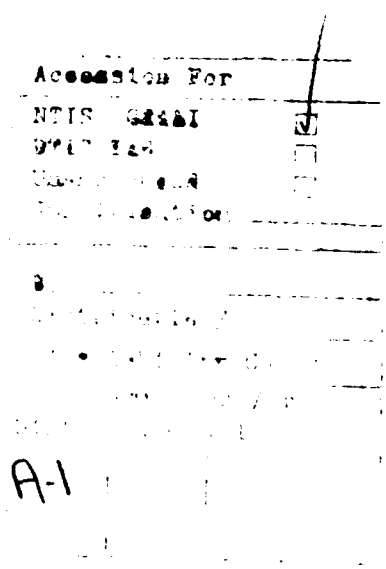
George B. Semb
John A. Ellis
John Araujo

Reviewed by
Nick Van Matre

Approved by
J. McLachlan

Released by
Thomas F. Finley
Captain, U. S. Navy
Commanding Officer
and
Richard C. Sorenson
Technical Director (Acting)

Approved for public release;
distribution is unlimited.



Navy Personnel Research and Development Center
San Diego, California 92152-6800

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATE COVERED
	August 1992	Final--Sep 89-Oct 91

4. TITLE AND SUBTITLE
Long Term Memory for Different Types of Classroom Knowledge

5. FUNDING NUMBERS
Program Element: 0602233N
Work Unit: RM33T23.05

6. AUTHOR(S)
George B. Semb, John A. Ellis, John Araujo

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
San Diego, California 92152-6800

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TR-92-18

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Chief of Naval Research (ONT-222)
800 North Quincy Street,
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Functional Area: Training Research
Product Line: Schoolhouse Training
Effort: Learning Processes

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Two studies examined students' long term retention of knowledge learned in college courses. In Experiment I retention for four different types of tasks was measured 4 and 11 months after the term ended. Overall students retained a great deal of what they originally learned. Retention for the recognition, recall, comprehension, and mental skill learning tasks differed only for the recall task which was significantly lower than the others. Experiment II analyzed knowledge retention in student tutors. Tutors retained more knowledge after four months than did the students they tutored. This suggests that tutoring, a type of overlearning, has positive effects that are maintained over time.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Long term memory, retention, learning task	21
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Foreword

This technical report is the second of two studies of knowledge retention that were conducted at the University of Kansas under the Manpower, Personnel and Training Technology (NP2A) Block of the 6.2 Mission Support Technology Program Element 0602233N, which was sponsored by the Chief of Naval Research (ONT-222). The goal of both studies was to determine how well information taught in school is retained. The research documented in this report builds on the first study by examining knowledge retention for different types of learning tasks. The theoretical background and methodology of the two studies are related, consequently, portions of the first two sections of this report are similar to the backgound and methodology sections of the first report, Semb, Ellis, and Montague (1990). However, this report describes both the theoretical and practical differences between the two studies.

The recommendations provided in this report are intended for use by the Chief of Naval Education and Training and the Chief of Naval Technical Training in determining instructional strategies for teaching knowledge in Navy classrooms. They are also to be used in an advanced development 6.3 project to develop and test instructional techniques for the classrooms of the future.

THOMAS F. FINLEY
Captain, U.S. Navy
Commanding Officer

RICHARD C. SORENSON
Technical Director (Acting)

Summary

Problem and Background

Military trainees frequently do not remember much of what they have been taught in school by the time they start working on the job. However, much of the research on memory loss has focused on procedural/psychomotor skills and tasks (e.g., performing preventive maintenance, operating equipment). While the research on skill retention has been extensive, there has not been much work on memory for knowledge taught in school.

Objectives

The objectives of this effort were (1) to determine if the amount of information retained from an introductory college course differs for different learning tasks (2) to determine the effects of serving as a course tutor on retention.

Method

Experiment I compared student performance after a 4- and an 11-month retention interval on four learning tasks: recognition, recall, comprehension, and mental skills. Because it was impossible to retest all the students who took the 4-month retention exam at the 11-month interval, the two retention intervals involved different numbers of students--the 11-month group (41 students) was a subset of the 4-month group (85 students). The primary performance measure was the score on the end-of-course exam composed of items that tested the four tasks. There were four forms of the exam and, at the 4-month interval, students were retested with the same form they had taken at the end of the course and were also tested with a different form. Students took the same two tests at the 11-month interval that they had taken at the 4-month interval.

Experiment II investigated the effect of serving as a course tutor on retention. Tutors were tested at the end of their tutoring semester and four months later. The comparison group was the students they tutored.

Results and Discussion

For Experiment I, a repeated measures analysis of variance (ANOVA) revealed significant decline in performance between the end-of-course test and the 4-month retention test, and between the 4-month and 11-month interval. Specifically, the results were that recall test items were not retained as well as recognition, comprehension, and mental skill items which were all retained equally well.

For Experiment II, both tutors and students scored near 90% (this score is at ceiling for the test) at the end of the course, but after four months tutors scored (79%) significantly higher than their former students (73%). The authors hypothesized that teaching the content resulted in additional opportunities for learning which in turn led to enhanced retention.

Recommendations

The Chief of Naval Education and Training and its subordinate commands should review their classroom testing programs in accordance with the testing guidelines provided in this report. The testing guidelines apply to knowledge taught in classrooms only and should not be used to review hands-on performance tests conducted in on-the-job or laboratory situations.

Contents

Introduction

Background

Military trainees frequently do not remember much of what they have been taught in school by the time they start working on the job (Hagman & Rose, 1983; Schendel, Shields, & Katz, 1978; Vineberg, 1975). Some of the factors that have been shown to contribute to this loss include type and complexity of the job-task, amount and quality of the initial training, length of time without practice, and amount and quality of experience encountered on-the-job (Farr, 1987; Montague, Wetzel, & Konoske, 1983). Most of the research investigating these factors has focused on procedural/psychomotor skills and tasks (e.g., performing preventive maintenance, operating equipment). After a number of years of research on these types of tasks, the Army integrated the findings and published a *User's Manual for Predicting Military Task Retention.* (1985) with an accompanying form, TRADOC Form 321-R.

While the research on retention of procedural/psychomotor skills has been extensive, memory for knowledge taught in school has received little attention (Neisser, 1982), even though memory for knowledge learned in school is an important issue. In the civilian world, the existence of high school and college education rests on the assumption that people remember something of what they learn. In the armed forces, knowledge retention is even more critical. Presumably, the knowledge that students learn in military classrooms is prerequisite for knowing when and how to perform jobs and tasks in the real world. Thus, it is important to know how much and what kind of knowledge students remember. The conventional wisdom is that most of what is learned in school is quickly forgotten. In a book on memory, Higbee (1977) states that "people . . . forget what they learned in school (usually within a short time after an exam)." Bahrick (1979) says that "much of the information acquired in classrooms is lost soon after final examinations are taken."

Until recently, few studies on retention of knowledge taught in schools have been published in the traditional educational and psychological research literature. In 1982, Neisser observed that "It is difficult to find even a single study, ancient or modern, of what is retained from academic instruction" (p.5). However, studies have been published in discipline-specific journals such as medical education, language instruction, civil engineering, and environmental science. A number of published articles also included retention measures as an after-thought. Semb and Ellis (1992) located and reviewed a number of these studies. Contrary to the conventional wisdom, these studies provide evidence that we remember much of what is taught in school. Semb and Ellis (1992) conclude that the amount of information acquired in classrooms that is lost over time is small (from 10% to 20% when measured by recognition tests).

Semb, Ellis, and Montague (1990) argue that, although the content of college courses and Navy apprentice training courses may be quite different, the strategies for teaching and learning that content are very much the same; both types of courses teach facts (e.g., names, dates, definitions), introductory concepts and principles, and, in some cases, problem solving skills. They conclude that examining what college students retain from an introductory college course should tell us something about what Navy personnel will retain from a Navy apprentice course.

In their study of retention from an introductory college course Semb, Ellis, and Montague (1990) varied the amount of original learning, the length of the retention interval, and the way in

which retention was tested. They found better retention for higher degrees of original learning, a decrease in performance over 4- and 11-month retention intervals, and better performance on a retention test that was identical to the test of original learning than on a test that covered the same content but with different questions. Overall, retention for the college course content was high; after 4-months students lost less than 10 percent on the identical retention test and less than 20 percent on the different test. This finding is consistent with the studies reviewed by Semb and Ellis (1992). Further, two studies of knowledge retained from Navy "A" School training, included in the Semb and Ellis (1992) review, support the argument that college courses and Navy apprentice training courses are comparable. Ellis (1980) found that "A" school students lost less than and Wetzel, Konoske, and Montague (1983) observed losses of less than 10 percent after 1-month when "A" school students were tested with an identical test.

Problem

The current studies replicate and extend the Semb, Ellis, and Montague (1990) research. The first experiment (Experiment I) examines retention for different learning tasks. The test items in Semb, Ellis, and Montague (1990) study were either multiple-choice or true-false and tested only the student's ability to recognize information. The dependent measures in Experiment I also test recognition memory but, in addition, they include test items that require recall of information, comprehension and problem solving (mental skills). Memory for these tasks, which are more cognitively demanding, is important for Navy personnel who operate complex systems in today's highly variable operational environment. Previous studies that have compared retention curves for recognition and recall measures have found recognition loss slower in both laboratory studies of list learning (e.g., Postman & Rau, 1957) and classroom studies of learning (e.g., Eurich 1934). In studies that compared retention of recognition, recall, and other types of knowledge tasks (e.g., application, interpretation, comprehension, problem solving), the results have not been as clear. Both Wert (1937) and Tyler (1933) found that interpretation and application skills increased over time while recognition and recall declined. In a study by Glasnapp, Poggio, and Ory (1978), which used Bloom's (1956) taxonomy to categorize items, there was a decline in retention for knowledge items but not for comprehension items. Also, application items in their study declined more rapidly than knowledge items. However, these differences were not statistically tested nor reported by Glasnapp et al. (1978). We derived these differences from treatment means presented in tables in their article. Halpin and Halpin (1982) found no differences in retention between conceptual and knowledge items when retention was measured by a short answer test. However, Wetzel, Konoske, and Montague (1983) found poorer retention for computational items than for knowledge items after a one 1-month interval in a Navy operator course.

One problem in many of these studies, which could have contributed to the conflicting results, was the failure to reliably define learning tasks and categorize test items for different learning requirements. Even the Glasnapp et.al. (1978) study, which used Bloom's taxonomy, did not report a systematic or reliable procedure for item classification.

The second experiment (Experiment II) compares student tutors who teach the course material with regular students to determine if serving as a tutor affects retention. The Navy has begun to experiment with peer tutoring in some schools (e.g., Naval Training Center, Great Lakes) and data on knowledge retention for tutors would provide more evidence on this process. It is expected that effects of tutoring will be similar to the effects of degree of original learning and amount of

overlearning as tutoring involves aspects of both of these retention related variables (Farr, 1987; Semb & Ellis, 1992). Tutors in the Experiment II first learned the material as students, and then re-learned and practiced it as tutors. This experience should result in enhanced retention similar to what was observed in early laboratory studies on overlearning (e.g. Krueger 1929).

Objectives

The objectives of this effort were (1) to determine if the amount of information retained from an introductory college course differs for different learning tasks and (2) to determine the effects on retention of serving as a course tutor.

Experiment I

Experiment I addresses the question: Does the amount of forgetting differ for different learning tasks? Answering this question requires defining learning tasks and reliably classifying the test items designed to measure the tasks. This requires at least two subject matter experts who are versed not only in course content but also in item and learning task definitions. Four types of learning tasks were defined and used to classify test items: recognition, recall, comprehension, and mental skills. Retention for each type of task was assessed after 4- and 11-month intervals. Based on previous studies, retention for recognition items is expected to be better than for recall items. Although, it is difficult to predict the amount of loss for comprehension and problem solving tasks, a conservative estimate would be that loss for these tasks would be less than for recall.

Method

Subjects

Subjects were students enrolled in 14 sections of a one-semester introductory child psychology course at the University of Kansas. Only students who were not psychology majors, who took the comprehensive final exam at the end of the semester, and who agreed to participate in the study by signing a University-approved research consent form and for whom American College Test (ACT) scores and University grade-point averages were available were considered for inclusion as subjects. Psychology majors were excluded because they were likely to take additional psychology courses during the retention interval. Students who were actually included as subjects took both forms of the retention exams at each of the two retention intervals. Because all the students who took the 4-month retention exam could not be retested at the 11-month interval, the two retention intervals involved different numbers of students--the 11-month group was a subset of the 4-month group. At 4 months, there were 85 students and at 11-months, there were 41 students.

Instructional Procedures

The course was divided into 12 content units, each of which covered about one chapter in the course text and accompanying study guide. Students were allowed to work at their own rate. All students took a 14-item quiz at the end of each unit and a 28-item midterm exam after every

fourth unit. The course ended with a comprehensive 72-item final exam. The 72-item course final exam was administered as a pretest at the beginning of the semester. It was one of four forms of the course final exam that was re-administered as the final at the end of the semester and, subsequently, as the retention test. Students had two chances to take each quiz, midterm, and end-of-course exam. The highest scores on each counted toward their grade in the course. Students received feedback from tutors and could ask the tutor for help.

Dependent Measures

The primary dependent measure was performance on the course exam, which was used as a pretest, an end-of-course final, and the retention tests. Each of the four forms of the exam, contained 60 four- or five-alternative multiple-choice items and 12 fill-in items. Each form contained six independent items from each unit. Items emphasized major concepts.

Item Categories. The four item categories were based on four learning tasks: recognition, recall, comprehension, and mental skills, and recall. Of the 60 multiple-choice items on the exam, 30 were recognition, 15 were comprehension, and 15 involved mental skills; the 12 fill-ins items were recall. Items were categorized using a learning task taxonomy similar to that described by Taylor and Ellis (1991). This classification scheme is based on earlier work by Merrill (1983), and Montague, Ellis, and Wulfeck (1983).

Recognition items required verbatim identification of terms, facts, definitions, concepts, and lists of characteristics taken directly from the text without paraphrasing or adding any new information. In the present study, they were four-or five-alternative multiple-choice questions.

Recall items involved remembering specific terms, facts, definitions, concepts, and lists of characteristics. To answer recall items, students had to produce an answer from memory rather than select it from a list of alternatives. Like recognition items, recall items were taken directly from the text without paraphrasing or adding any new information. The stem of each recall item was identical to the stem of its companion recognition item; the items differed only in absence or presence of a list of alternatives.

Comprehension items required students to know the meaning of what was taught rather than to memorize as required to answer recognition items (Taylor & Ellis, 1991). This involved paraphrasing material, either in the stem or in the alternatives of the item. It could also involve comparing the alternatives and then selecting the best answer. Comprehension items were also of a multiple-choice format. The items were developed following guidelines described in Anderson (1972).

Mental skill items emphasized the student's ability to use knowledge in a situation not explicitly taught in the course texts. Novelty was a key feature of many mental skills items. For example, a question might present a novel example, not presented in the text, and ask students to identify the concept. Another dimension of novelty involved having one or more untaught comparisons among alternatives. Such alternatives were plausible but not related to course content. Mental skill items were also multiple-choice.

Test Construction. Items were categorized by the course developer and two instructors, each of whom had taught the course at least three years. These individuals were experienced with both course content and item-category definitions. First, the course developer assigned items to each item category that covered major concepts. Next, each instructor categorized items and inspected them for any overlap in the specific content they tested. This inspection was particularly critical for recognition and recall items because companion items covered identical content. Ten percent of the items were discarded because they covered the same concepts. For the remaining items raters agreed on item category assignment 83.1 percent of the time. Items they could not agree on were discarded.

Questions were then randomly assigned without replacement to the four forms of the exam with the following constraints: (1) no question appeared on more than one form of the exam, (2) no more than 36 items on any form appeared elsewhere (study guide, quizzes, or midterms), and (3) no exam contained more than one question covering any single major concept. To determine the equivalence of the four forms, scores achieved by 80 students who did not participate in the study were randomly sampled (20 for each of the 4 forms). The means were 74.2, 75.2, 68.8, and 71.4 percent correct. There were no significant differences among forms.

Testing Procedures

Retention Tests. The first set of retention tests was administered 4 months after the course ended. Students who met the eligibility criteria were contacted by telephone and invited to take the retention exams. As an incentive to participate, they were offered a 1 in 10 chance to win $15 and a 1 in 100 chance to win $100. All students who participated took two retention exams, one they had not seen before (different) and the identical form on which they had scored highest during the previous semester (same).

The second set of retention exams was administered 11 months after the course ended. Students were contacted by telephone and invited to take the retention exams. As an incentive, they were offered $10 to retake the exams, plus a 1 in 50 chance to win $100. They took the same two retention exams that they had taken at the end of 4 months.

Order of Exams.A student could potentially take all four forms of the course exam: pretest, two finals, and the different retention test. Students who took the final only once were exposed to three forms; students who took it twice were exposed to all four. Exam order was determined by assigning students randomly to one of the 24 possible exam order sequences.

Results and Discussion

Overall Retention and Order Effects

Overall, students retained over 75 percent of what they originally learned after 4 months and over 70 percent after 11 months. These figures are lower than those reported by Semb, Ellis and Montague (1990). This effect is attributable to differences among item types. Performance scores were lower for the fill-in-the-blank items, which constituted 16 percent of the items in Experiment I. Table 1 presents the overall performance data for the 4-month and 11-month groups.

Table 1

Means and Standard Deviations for Overall Percent Correct Responses for the 4-month and 11-month Groups

Groups	Mean	SD
4-month Group		
End-of-course Exam	71.9	10.7
Retention Exam, Same Form	56.6	12.1
Retention Exam, Different Form	52.4	10.4
11-month Group		
End-of-course Exam	73.5	9.0
Retention exam, Same Form	53.9	9.0
Retention Exam, Different Form	51.1	8.3

Similar to the results of the Semb, Ellis, and Montague (1990) study, the order of presentation of the retention exams (same-different versus different-same) did not affect performance, but students performed better on the same exam, regardless of whether it was taken first or second.

Item Categories (Learning Tasks)

A one-way analysis of variance (ANOVA) was performed on the gain scores from the pretest to the end-of-course test. Gains among categories were significantly different, $F(3, 252) = 21.7$, $MSE = 307.9$, $p < 0.01$. A post-hoc analysis revealed that gains for recognition, comprehension, and mental skills were significantly ($p < 0.01$) lower than for recall items. This is because accuracy on recall (fill-in-the-blank) items on the pretest was extremely low (9.3%) compared to recognition (45.4%), comprehension (41.9%), and mental skills (39.2%) items which were all multiple-choice items.

A 2 (same vs. different) by 4 (item category) ANOVA was performed on the loss scores at the 4-month retention interval. Performance scores on the same form of the exam were significantly higher, $F(1,84) = 9.97$, $MSE = 279.58$, $p < 0.01$; there was a significant effect for item category, $F(3,252) = 12.33$, $MSE = 408.24$, $p < 0.01$, but no significant interaction. The only significant item category difference was that performance scores on recall items were lower than performance on the other three item types. A similar analysis was performed on loss scores at the 11-month retention interval. Retention interval was added as a factor to create a 2 (same vs. different) by 4 (item category) by 2(4-month and 11-month interval) repeated measures design. The findings were identical to those at the 4-month interval with the addition of a significant main effect for retention interval $F(1,40) = 36.76$, $MSE = 115.41$, $p < 0.01$; performance was lower at 11-months than at 4-months. There were no significant interactions. Table 2 presents the mean percent loss scores and standard deviations at each of the two observation periods.

Table 2

Mean Percent Loss Scores and Standard Deviations for the Same and Different Exam forms at 4-months and 11-months

	4-month Group		11-month Group	
Item Category	Mean	SD	Mean	SD
Same Form Loss (End-of-Course to End-of-Interval)				
Recognition	-13.7	12.7	-18.7	11.7
Recall	-25.0	20.6	-28.3	19.2
Comprehension	-12.7	16.6	-15.3	17.1
Mental Skills	-13.7	18.1	-18.7	15.8
Different Form Loss (End-of-Course to End-of-Interval)				
Recognition	-16.9	14.2	-20.2	10.8
Recall	-27.9	25.4	-36.2	20.8
Comprehension	-18.6	18.9	-16.9	18.8
Mental Skills	-17.9	21.0	-21.3	23.4

The findings for recognition and recall tasks are consistent with those of Bahrick (1984), Eurich (1934), and Halpin and Halpin (1982), who reported that recognition items are retained better than recall items. The failure to find differences among recognition, comprehension, and mental skills tasks, however, is not consistent with the Glasnapp et al. (1978), Tyler (1933), and Wert (1937), studies. Unlike the present investigation, none of these studies clearly defined learning tasks and reliably classified test items.

Experiment II

The teaching staff in Experiment I included students who had taken the course previously and had returned to serve as tutors. This provided an opportunity to examine the effect that teaching the course material had on retention for these students. Farmer, Lachter, Blaustein, and Cole (1972) have shown that some tutoring benefits tutor performance, and Gaynor and Wolking (1974) have shown that peer tutors benefit from the experience. Several studies (Fitch, 1991; Gaynor, 1975; Johnson & Sulzer-Azaroff, 1978; Robin & Heselton, 1977; Sulzer-Azaroff, Johnson, Dean, & Freyman, 1977) have analyzed variables associated with tutor training and classroom performance. However, only Arp and Semb (1977) analyzed tutors' academic performance. They found that tutors gained eight percent more from the pretest to the end-of-the-course exam than the students they tutored, but these results must be interpreted with caution because they are based on observations of only four individuals. At present, no other data are available indicating what tutors retain from their tutoring experience. Thus, Experiment II addressed is whether tutoring has the same effect as overlearning and promotes higher levels of retention for tutors than for students who only complete the course.

Method

Subjects and Setting

The subjects were tutors in the introductory child development course described in Experiment I. The tutors were students who had taken the course during the previous semester and who were selected by the teaching staff to serve as course assistants. Of the 64 tutors who took the end-of-course exam, 63 agreed to participate in the study by signing a university-approved research consent form, but only 34 had ACT test scores and grade-point averages available. Of these, 25 took the retention exam, and it is on these 25 students that the present analysis is based. The comparison subjects were 44 students who participated in the Semb, Ellis, and Montague (1990) study.

Dependent Measures and Testing Procedures

The dependent measures were scores on the end-of-course exam and retention exams, described in the Semb, Ellis, and Montague (1990) study.

Tutors took the course exam for the first time at the end of the semester in which they were enrolled in the course as students. The end-of-course exam counted for up to 30 percent of their grade in the course. They took an alternate form of the exam 4 months later at the beginning of the semester in which they tutored. They were encouraged to answer every item and were told that their performance would be used to determine what material they would have to study before they could tutor students. At the end of the tutoring semester, they took another form of the exam, which was unannounced and without any consequences. The retention test was administered 4 months later. Tutors were contacted by telephone and invited to take the retention exam. As an incentive, they were offered $10 to take the exam. Exam order sequences for each tutor were determined randomly across the four testing times. For comparison purposes, the data used in this experiment were from the end-of-course exam and the 4-month different form exam.

Results

Demographic Characteristics

The 25 tutors had a mean composite ACT score of 21.6, a mean GPA of 2.91, and had completed 2.9 years of university education. The 44 students had a mean composite ACT score of 20.7, a mean GPA of 2.77, and had completed a mean of 2.3 years of university education.

Academic Performance

Table 3 presents percent correct responses for tutors and students. A one-way ANOVA revealed significant differences across the four testing times for tutors, $F(3,72) = 30.9$, $p < .01$, $MSE = 9.64$. Post hoc comparisons revealed that (1) tutors lost 8.9 percent in knowledge between the end of the semester in which they were enrolled as students and the beginning of the semester when they tutored, $t(24) = 6.67$, $p < .01$, (2) by the time tutoring ended, they had gained back 8.3 percent $t(24) = 6.82$, $p < .01$, and (3) after four months, they lost 7.6 percent $t(24) = 8.16$, $p < .01$.

Table 3

Means and Standard Deviations for Percent Correct Responses for Tutors and Students

	Treatment Group			
	Tutors		Students	
Test Administrations	Mean	SD	Mean	SD
End of Student Semester-Tutors Only	87.1	3.8	NA	NA
Begin Tutoring-Tutors Only	78.2	6.1	NA	NA
End of Tutoring-Tutors and Students	86.5	4.1	87.3	6.2
4-month Retention-Tutors and Students	78.9	5.6	73.5	5.7

A 2 by 2 ANOVA compared groups (tutors and students) at the two testing times (end of the course and 4-month retention interval). There was no significant between group effect; however, testing time and the group-by-time interaction were both significant, $F(1,67) = 2.96$, $p < .01$, MSE = 12.4 and $F(1,67) = 24.2$, $p < .01$, MSE = 12.4, respectively. At the end of the course, the students performed at a level comparable to their tutors, but, after 4 months, tutors lost 7.6 percent while students lost 13.8 percent (see Table 3).

Discussion

The results indicate that tutors retain much of what they practice in the classroom and that they retain significantly more over time than do the students they tutor. These results suggest that tutoring has academic benefits in addition to social and personal benefits (Keller, 1968). The academic effects may be due to several factors including increased and continued exposure to course content and the experience of tutoring students.

Tutors did not perform any better than their students at the end of the semester even though they had considerably more exposure to the course material. However, the final exam counted for 30 percent of the student's course grade while the tutors' performance had no consequences. Furthermore, both tutors and students achieved nearly 90 percent correct; based on past exam statistics performance cannot go much higher. High levels of performance which are typical of Personalized Systems of Instruction (PSI) classes (Hursh, 1976; Kulik, Jaska, Kulik, 1978; Kulik, Kulik, & Bangert-Drowns, 1990), result in ceiling effects that may mask differences.

Discussion, Conclusions, and Summary

The results of both experiments suggest that students retain a great deal (over 80% for recognition items) of what they learn in college courses, contrary to popular belief. Both experiments also provide information on how several variables influence knowledge learned in the classroom (Farr, 1987; Semb & Ellis, 1992). The results of Experiment I show that the type of learning task is an important determinant of retention. Specifically, recognition, comprehension,

and mental skill tasks were retained better across both retention intervals than were recall tasks. Further, performance on recognition, comprehension, and mental skill tasks was equivalent. This finding is not consistent with some previous studies (e.g., Glasnapp et al., 1978) and deserves further examination. Also, in Experiment I, retention was greater after 4 months than after 11 months. While this result was expected, the decreases in performance between the 4- and 11-month retention intervals were small (less than 10% for any dependent measure).

In Experiment II, tutoring provided the additional opportunities for learning, which facilitated long-term retention. This finding is consistent with laboratory studies of overlearning and studies in which the degree of original learning was manipulated (Krueger, 1929).

Guidelines for Knowledge Test Design and Development

The following guidelines should be used only for knowledge taught in classrooms and not for hands-on performance tests conducted in on-the-job or laboratory situations.

Recognition level course tests (e.g., multiple-choice, matching, true-false) are appropriate when the overall course training goal is to prepare students for heavily supervised on-the-job training (OJT) by familiarizing them with job terminology, definitions, technical documentation, job duties, and other factual information required for job performance. Based on the findings reported in this study if students can pass recognition tests with scores of 10 percent over the minimum passing score (typically 60 or 70 percent) they should be capable of passing identical tests following retention intervals of up to 1 year. Therefore, a passing score of 10 percent over the minimum score should adequately prepare course graduates for their subsequent duty assignments. Delays of over 2-months will occur between graduation and job assignment, schools should ensure that their students exceed the minimum passing score by 10 percent.

Course tests that require recall of information taught in classrooms are appropriate when the overall course training goals are to prepare students for minimally supervised OJT or unsupervised job performance. For these job conditions students need to be more than just familiar with the knowledge required for job performance. The results of this study and other studies that have investigated the ability to recall information after a long retention interval indicate that levels of original learning for knowledge that must be recalled should be high to ensure subsequent recall accuracy. Based on the current findings, passing scores for recall tests should be set at over 90 percent to ensure that performance will be above 60 percent following a 4-month or 11-month interval. Significantly longer training times to reach this higher level of achievement should be expected.

For courses that prepare students for jobs that require the application of complex mental skills (e.g., classification, problem solving, diagnosis, troubleshooting) multiple-choice, matching, short-answer, essay, fill-in, true-false, and performance test items are appropriate. Unfortunately, the retention data for these types of tasks are inconclusive. The findings from the present study show small amounts of loss which is similar to the data for the retention of factual knowledge. However, other studies have observed more substantial loss patterns (e.g., Wetzel, Konoske, & Montague, 1983, 1984). Further research is required to clarify and reconcile these discrepancies.

Functionally, peer tutoring behaves like overlearning in increasing resistance to forgetting. When feasible, peer tutoring should be implemented in Navy classrooms both to facilitate original learning and to enhance retention.

Recommendations

The Chief of Naval Education and Training and its subordinate commands should review their classroom testing programs in accordance with the testing guidelines provided in this report.

References

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42*, 145-170.

Arp, L., & Semb, G. (1977). An analysis of the use of student proctors in a personalized college business course. *Journal of Personalized Instruction, 2*, 92-95.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*, 296-308.

Bahrick, H. P. (1984). Memory and people. In J. Harris (Ed.), *Everyday memory: Actions and absent-mindedness*. New York: Academic Press.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives.* New York: David McKay Company, Inc.

Ellis, J. A. (1980). *Long-term retention of factual information* (NPRDC-TN 80-5). San Diego: Navy Personnel Research and Development Center.

Eurich, A. D. (1934). Retention of knowledge acquired in a course in general psychology. *Journal of Applied Psychology, 18*, 209-219.

Farmer, J., Lachter, G. D., Blaustein, J. J., & Cole, B. K. (1972). The role of proctoring in personalized instruction. *Journal of Applied Behavior Analysis, 5*, 401-404.

Farr, M. J. (1987). *The long-term retention of knowledge and skills: A cognitive and instructional perspective*. New York: Springer-Verlag.

Fitch, M. A. (1991, April). *Peer teacher training as a function of role playing and performance evaluation tasks: Effects on student course performance and satisfaction.* Paper presented at the annual meeting of the American Research Association, Chicago.

Gaynor, J. F. (1975). Analysis of the proctoring variable in PSI courses. *Educational Psychology, 15*, 30-35.

Gaynor, J. F., & Wolking, W. D. (1974). The effectiveness of currently enrolled student proctors in an undergraduate special education course. *Journal of Applied Behavior Analysis, 7*, 263-269.

Glasnapp, D. R., Poggio, J. P., & Ory, J. C. (1978). End-of-course and long-term retention outcomes for mastery and nonmastery learning paradigms. *5*, 595-603.

Hagman, J. D., & Rose A. M. (1983). Retention of military tasks: A review. *Human Factors, 25*, 199-213.

Halpin, G., & Halpin G. (1982). Experimental investigation of the effects of study and testing on student learning, retention, and ratings of instruction. *Journal of Educational Psychology, 74*, 32-38.

Higbee, K. L. (1977). *Your memory: How it works and how to improve it.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Hursh, D. E. (1976). Personalized *systems of instruction: What do the data indicate?* Journal of Personalized Instruction, 1, 91-105.

Johnson, K. R., & Sulzer-Azaroff, B. (1978). An experimental analysis of proctor prompting behavior in a personalized instruction course. *Journal of Personalized Instruction, 3*, 122-130.

Keller, F. S. (1968). Good-bye, teacher ... . *Journal of Applied Behavior Analysis, 1*, 79-89.

Krueger, W. C. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71-78.

Kulik, J. A., Jaska, P., & Kulik, C. C. (1978). Research on component features of Keller's personalized system of instruction. *Journal of Personalized Instruction, 3*, 2-14.

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*, 265-299.

Merrill, M. D. (1983). Component display theory. In C. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status.* Hillsdale, NJ: Lawrence Earlbaum. 887-892.

Montague, W. E., Ellis, J. A., & Wulfeck, W. H. (1983, June). Instructional quality inventory: A formative evaluation tool for instructional development. *Performance and Instruction Journal,* 11-14.

Montague, W E., Wetzel, S. K., & Konoske, P. (1983). *A tool to help diagnose performance problems: The memory maintenance quality inventory (MMQI).* Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA.

Neisser, U. (1982). *Memory observed: Remembering in natural contexts.* New York: Freeman.

Postman, L. W., & Rau, L. (1957). Retention as a function of method of measurement. University of California. *Publications in Psychology,* 8, 217-270.

Robin, A. L., & Heselton, P. (1977). Proctor training: The effects of a manual versus direct training. *Journal of Personalized Instruction, 2,* 19-24.

Schendel, J., Shields, J., & Katz, M. (1978). *Retention of motor skills: Review.* (Technical Paper 313.) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Semb, G., & Ellis, J. A. (1992). *Knowledge learned in college: What is remembered?* Manuscript submitted for publication.

Semb, G., Ellis, J. A., & Montague, W. E. (1990). *Long term memory for knowledge taught in school.* (NPRDC-TR 91-1). San Diego: Navy Personnel Research and Development Center. (AD-A228 791).

Sulzer-Azaroff, B., Johnson, K. R., Dean, M. R., & Freyman, D. R. (1977). An experimental analysis of proctor quiz-scoring accuracy in personalized instruction courses. *Journal of Personalized Instruction, 2*, 143-149.

Taylor, B., & Ellis, J. (1991). An evaluation of instructional systems development in the Navy. *Educational Technology Research and Development, 39*, 1, 93-103.

Tyler, R. W. (1933) Permanence of learning. *Journal of Higher Education, 4*, 203-204.

User's manual for predicting military task retention. (1985). Fort Eustis, VA: Army Training Board.

Vineberg, R. (1975). *A study of the retention of skills and knowledge in basic training.* (HumRRO TR 75-10). Alexandria, VA: Human Resources Research Organization.

Wert, J. E. (1937). Twin examination assumptions. *Journal of Higher Education, 8*, 136-140.

Wetzel, S. K., Konoske, P. J., & Montague, W. E. (1983). *Estimating skill degradation for aviation antisubmarine warfare operators (AWs): Loss of skill and knowledge following training.* (NPRDC-TR-83-31). San Diego: Navy Personnel Research and Development Center. (AD-A129 407).

Wetzel, S. K., Konoske, P. J., & Montague, W. E. (1984). *Estimating skill loss throughout a Navy technical training pipeline.* (NPRDC-TR-84-7). San Diego: Navy Personnel Research and Development Center. (AD-A136 636).

Distribution List

Distribution:
Office of Chief of Naval Research (ONT-222), (OCNR-10), (ONT-20)
Defense Technical Information Center (DTIC) (4)

Copy to:
Chief of Naval Education and Training (CNET-00), (CNET-N2), (CNET-N5), (CNET-N54), (CNET-L01) (6), (NETPMSA-03) (2), (NETPMSA-04) (2), (CNET-01), (CNET-N1), (CNET-N3), (CNET-N12)
Chief of Naval Technical Training (Code 00)
Commander, Naval Training Center, Great Lakes, IL
Commanding Officer, Service School Command, Great Lakes, IL
Commander, Naval Training Center, San Diego, CA
Commander, Naval Training Center, Orlando, FL
Commanding Officer, Service School Command, San Diego, CA (Code 3200)
Commanding Officer, Service School Command, Naval Training Center, Orlando, FL
Commander, Training Command, U.S. Atlantic Fleet
Commander, Training Command, U.S. Atlantic Fleet (Code 01A)
Commander, Training Command, U.S. Pacific Fleet (Code N-31)
Office of Chief of Naval Research (OCNR-1142CS)
Technical Director, U.S. ARI, Behavioral and Social Sciences, Alexandria, VA (PERI-ZT)
Director, Training Systems Development, Randolph Air Force Base, TX (HQ ATC/XPRS)
Institute for Defense Analyses, Science and Technology Division
Pentagon Library
Commanding Officer, Fleet Anti-Submarine Warfare Training Center, Atlantic (N-12)
Bureau of Naval Personnel (PERS-05), (PERS-116),
Commanding Officer, Sea-Based Weapons and Advanced Tactics School, Pacific
U.S. ARI, Behavioral and Social Sciences, Alexandria, VA
Department of the Air Force, DET 5, Armstrong Laboratory Directorate, Brooks Air Force Base, TX
AL/DOKLO Technical Library, Brooks Air Force Base, TX
Director of Research, U.S. Naval Academy
Center for Naval Analyses, Acquisitions Unit
Center for Naval Analyses