NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER, SAN DIEGO, CALIFORNIA 92152

NPRDC TR 78-15 MARCH 1978

IMMEDIATE VS. DELAYED FEEDBACK IN A COMPUTER-MANAGED TEST: EFFECTS ON LONG-TERM RETENTION

NPRDC TR 78-15 March 1978

IMMEDIATE VS. DELAYED FEEDBACK IN A COMPUTER-MANAGED TEST: EFFECTS ON LONG-TERM RETENTION

Persis T. Sturges
California State University
Chico, California 95929

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Navy Personnel Research and Development Center under Contract N00123-76-C-1066.

Reviewed by
William E. Montague

Approved by
James J. Regan
Technical Director

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPRDC TR 78-15

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): IMMEDIATE VS. DELAYED FEEDBACK IN A COMPUTER-MANAGED TEST: EFFECTS ON LONG-TERM RETENTION

5. TYPE OF REPORT & PERIOD COVERED: Technical Rep[redacted]t, Mar[redacted]76-Aug[redacted]76

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Persis T. Sturges

8. CONTRACT OR GRANT NUMBER(s): N00123-76-C-1066

9. PERFORMING ORGANIZATION NAME AND ADDRESS: California State University, Chico, California 95929

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62709E

11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, Arlington, Virginia 22209

12. REPORT DATE: February 1978

13. NUMBER OF PAGES: 35

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Navy Personnel Research and Development Center, San Diego, California 92152 (Code 304)

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Feedback
Learning
Instruction
Computer-based Instruction
Education
Retention

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Four groups of college undergraduates took a multiple-choice computer-managed test. Three of these groups received informative feedback (the entire item with the correct answer identified) either (1) immediately item-by-item (2-second delay), (2) following the entire test (20-minute delay), or (3) 24 hours later (24-hour delay). The fourth (control) group received no feedback. Scores on a criterion test, given 1 to 3 weeks later, showed that retention was significantly better for the two delayed feedback groups (20-minute and

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

24-hour delay) than for the immediate feedback group (2-second delay). These results confirmed previous findings of laboratory experiments--that retention following delayed feedback is not degraded by the delay.

FOREWORD

This research and development was sponsored by the Advanced Research Projects Agency under ARPA Order No. 3181, and is part of the ARPA training technology program. A principal objective of this program is the development of computer-based training technology for DoD-wide application.

This study is the first in a series investigating the relative effectiveness of providing immediate or delayed feedback for answers on computer-managed tests. The results provide important background information for potential application in Navy training whenever computer-managed testing is utilized. In contrast to the commonly held view that immediate feedback is essential to promote learning and retention, the results of this practical study confirmed the findings of prior laboratory studies--that retention following delayed feedback is not degraded by the delay.

Dr. William E. Montague served as the technical contract monitor.

J. J. CLARKIN
Commanding Officer

SUMMARY

Problem

Most programs for computer-managed and computer-assisted instruction are based on the principle that providing immediate informative feedback--that is, information about the correctness of the student's response--is essential if learning is to occur. However, findings of several experimental studies raise doubts as to whether it is necessary to provide immediate feedback to promote better retention or whether delays in informative feedback (i.e., when there is an interval between the student's response and the presentation of feedback, can be tolerated. Since it is extremely expensive to provide for immediate feedback in the design of computer-managed and computer-assisted instructional systems, sound data are needed to determine the relative effectiveness of the two methods.

Purpose

The purpose of the present experiment was to determine whether the findings of short laboratory studies conducted previously would generalize to computer-managed testing in a college course. The specific purpose was to determine whether delivering feedback immediately or after a delay interval differentially affected later retention.

Approach

As part of an undergraduate course, four groups of students were administered a computer-managed test comprised of 30 multiple-choice items. Three of the groups received informative feedback: the first, item-by-item, immediately after students had completed each item (2-second delay); the second, immediately after they had completed the entire set of items (20-minute delay); and the third, 24 hours later (24-hour delay). The fourth (control) group received no feedback. From 1 to 3 weeks later, a criterion test over the same material was administered to all groups. The criterion test included 47 multiple-choice items (30 that were the same as those in the initial test, and 17 others on the same material) and 10 short-answer items (five that were adapted from multiple-choice items in the initial test, and five others on the same material). For multiple-choice items in both tests, students recorded their chosen alternative and their confidence in the correctness of that choice. Also, an anxiety scale was administered to all groups before and after they took the computer-managed test, when they returned to the testing room 24 hours later to receive their test score, and after they took the criterion test; and to the feedback groups immediately after they had received and studied their feedback.

Analyses were conducted to determine whether there were any differences between the feedback groups in (1) performance on multiple-choice items included in both tests, multiple-choice items included in the criterion test only, and short-answer "same" and "different" items; (2) time required to answer the items and to study informative feedback; and (3) anxiety experienced before and after testing.

Findings

1. Analyses of performance on multiple-choice items that were included in both tests showed that feedback conditions had a significant effect on several measures:

a. The overall mean correct for the immediate feedback condition (2-second delay) and the two delayed feedback conditions (20-minute and 24-hour delay) combined was significantly greater than that for the no feedback (control) condition.

b. The mean correct for the two delayed feedback conditions combined was significantly greater than that for the immediate feedback condition. These results can be attributed to the feedback effects for items that were incorrect on the first tests; that is, for the proportion of items that were wrong on the first test and right on the second, reliable effects of feedback conditions were found, while for items that were correct on both tests, no effects were found.

c. Feedback conditions also had a significant effect on the amount of change in confidence ratings for items that were wrong on the initial test and right on the second. The 24-hour delay group had a significantly greater change in confidence ratings than any of the other groups; and the 20-minute delay group, a significantly greater change than the 2-second immediate feedback group.

d. When items were categorized in terms of their difficulty, feedback conditions had a significant effect only for the most difficult ones. The relationships among feedback conditions were the same as those reported in b above.

2. Feedback conditions had no significant effect on any of the other measures.

Conclusions

A consistent finding of previous laboratory experiments--that long-term retention following immediate feedback is <u>not</u> superior to that with delayed feedback--has been confirmed with computer-managed tests in an educational setting. In fact, some delay in presentation of feedback results in superior retention; and the longer 24-hour delay has a greater effect on the change in student's confidence in their answers.

Recommendations

1. Further research should be conducted to extend the findings of the present study by comparing the relative effects of immediate and delayed feedback under other experimental conditions (e.g., using different forms of feedback presentation and/or criterion test items and conducting repeated computer-managed tests with informative feedback throughout a course).

2. It is assumed that these results are due to an increase in student concentration on feedback that influences the level or breadth of processing of the remembered information and the feedback. Therefore, procedures that foster the breadth of processing should be developed and evaluated.

3. The appropriateness of requiring immediate feedback to be provided in curriculum development (NAVEDTRA 106A, 1975) should be reconsidered, since such conditions may not be optimum for learning.

CONTENTS

CONTENTS

INTRODUCTION

Problem

Most programs for programmed, computer-managed, and computer-assisted instruction are based on the principle that providing immediate informative feedback--that is, immediate information to the student about the correctness of his or her response to questions--is essential to promote learning and retention. However, findings of several laboratory experimental studies--that retention following delayed feedback (i.e., when there is an interval between the student's response and the presentation of feedback) is not degraded by the delay--raise doubts as to the validity of that principle. Since it is extremely laborious and expensive to provide for immediate feedback in the design of computer-managed and computer-assisted instructional systems, sound data are needed to determine whether these laboratory results generalize to actual course conditions.

Background

An extensive experimental literature exists that is concerned with corrective feedback and its effects on student learning and retention (Kulhavy, 1976). Concern for the effects of the delay of feedback resulted from: (1) the idea that since feedback was a form of reinforcement, it should function as food does in shaping the behavior of a hungry organism, and (2) the findings that delays in food reinforcement were often disruptive to performance. Therefore, if feedback reinforces correct responses, any delay should produce poorer learning and memory. The experimental studies in this area were conducted mostly in laboratory situations using a wide variety of learning tasks, ranging from simple discrimination studies for children (e.g., when learning to choose between geometric shapes) to segments of actual lesson materials for children and adults. Since a variety of procedures were used in these studies, terminological confusions arise in trying to summarize them and to extrapolate their results.

Delays in giving feedback about the correctness of answers to each test question can be introduced in various ways. For example, "immediate" feedback may be provided within a second or two after a student makes a response, while "delayed" feedback may be provided 10 or 20 seconds after the response. In other instances, the student might have to answer all questions on a test before any informative feedback is given. Thus, even if feedback is given "immediately" after the entire test, the specific feedback for the initial question is actually delayed by the length of time the student needs to answer the remaining questions and to receive feedback about all answers.

The effect of any feedback manipulation is evaluated by giving another test, a retention/criterion test, either a short time (seconds or minutes), or a longer time (usually a day or more) after feedback treatment is completed. The findings of studies providing short post-item-answer delays of less than a minute, and those providing longer, post-test delays on short retention intervals are consistent: Only occasionally does delay produce any detrimental effect on short-retention-interval test performance; in those cases, it usually involves the learning of simpler tasks. Most studies reveal no significant effect of delay. However, when retention tests are delayed 1 to 7 days, performance of subjects receiving feedback after a delay, is often superior to that of subjects receiving feedback immediately. It is important to note that no study found that <u>long-term</u> retention following immediate feedback was superior to that following delayed feedback.

Since the studies using simple learning tasks (e.g., Brackbill, 1964) differ in procedures and purpose from those concerned with tasks more relevant to instructional problems, for purposes of this report they will be ignored in favor of the more relevant studies. In the period since 1960, 13 experiments have been reported, all of which used academic-type materials and compared immediate feedback with intervals ranging from 2 seconds to 20 minutes with delayed feedback with intervals ranging from 24 to 48 hours. Thus, the following discussion will be limited to those 13 experiments, which were reported by English and Kinzer (1966); Kulhavy and Anderson (1972); More (1969); Newman, Williams, and Hiller (1974); Phye and Baller (1970); Sassenrath and Yonge (1968); Sturges (1969 and 1972--Experiments 1 and 2); Sturges and Crawford (1964--Experiments 1, 2, and 4); and Surber and Anderson (1975).

Delay Retention Effect

In 11 of these experiments, all but those reported by Newman et al. and Sturgis and Crawford (Experiment 4), it was found that retention following delayed feedback was superior to that following immediate feedback. This phenomenon has been called the "delay retention effect." Newman et al. had no evidence that students had learned from informative feedback, since the performance of the two feedback groups did not differ from that of a control group receiving no feedback. The feedback provided by Sturges and Crawford in their fourth experiment consisted of giving all alternatives, along with a cue directing the student to the correct answer.

The studies in which the delay retention effect occurred were identical in regard to the following conditions:

1. As indicated previously, the learning task concerned academic-type material.

2. Initial test items and the informative feedback were in a multiple-choice format.

3. Informative feedback was presented only once.

4. The items in the initial and retention tests were identical.

These studies also had a number of varied conditions:

1. In some studies, the initial test and presentation of feedback represented the student's first exposure to the learning material; in others, the student had studied the material before being tested.

2. In some studies, students were given a retention test immediately after receiving feedback and a long-term retention test later; in others, they were given only a long-term retention test.

3. In some studies, immediate feedback referred to item-by-item presentation immediately after the student made each response (approximately 2-second interval); in others, it referred to feedback presented after the student had responded to the entire set of test questions (20-minute interval). For the purposes of this report, the latter (20-minute interval) will be considered delayed feedback.

4. In one study (Phye & Baller, 1970), feedback was presented auditorily; in the rest, it was presented in printed form.

5. In one study (Sturges, 1972), retention was measured by a recall test as well as a multiple-choice retention test; in the rest, no recall test was included.

Variables Affecting Delay Retention Effect

The delay retention effect is affected by:

1. The form of informative feedback or the amount of information or stimulus aspects presented at feedback.

2. The accuracy of the student's initial response to an item.

In addition, a variable of potential importance is the amount of anxiety the student is experiencing when he receives feedback.

These variables are discussed in the following paragraphs.

Form of Informative Feedback. Sturges (1969) and Phye and Baller (1970) found that the delay retention effect occurred when feedback consisted of incorrect alternatives as well as the correct alternative but not when it consisted of only the correct alternative. Sturges (1972) extended this finding, when she found that the delay retention effect occurred with this form of information feedback only when the incorrect alternatives were relevant to the material in the retention test. The feedback provided by Sassenrath and Yonge (1968) also consisted of correct and incorrect alternatives; however, in some cases, it included the stem, while in others it did not. The delay retention effect occurred under both conditions.

Sturges (1972) also compared the effect of providing delayed feedback for six different types of feedback, including that used by Sturges and Crawford (1964) in their fourth experiment: all alternatives along with a cue directing the student to the correct answer. As indicated previously, using this method, retention after immediate feedback was equal to that with delayed feedback. On the basis of this finding, Sturges concluded information included in feedback determines how students respond to and thus learn from that feedback. She also concluded that the student's reaction to and amount learned from that feedback is affected by the length of the delay interval. That is, when immediate feedback is presented, the student can determine the correctness of his initial response by merely checking to see whether the number of that response agrees with the correct alternative. On the other hand, with delayed feedback, he must read the item and study all of the information included in feedback to determine whether his answer was correct. Retention test performance will be improved, then, when the student is required (at feedback) to recall and "think about" the information relevant to the requirements of the retention test.

Accuracy of Student's Initial Response. Kulhavy and Anderson (1972) suggested a "perseveration-interference" hypothesis, which may explain the delay retention effect. According to this hypothesis, for a time after making a response, the memory for the response "perseverates" temporarily. Thus, if the response is incorrect, this perseveration interferes with correcting the response when feedback is immediate. With delayed feedback, perseveration is over and, thus, there is no interference. Surber and Anderson (1975) gave support to this hypothesis, when they found that delayed feedback is more effective for items that were initially incorrect.

Anxiety Experienced at Feedback. According to the Drive Theory (Spence, 1958; Taylor, 1956), subjects experiencing a high degree of anxiety do not perform as well as those with a lower degree of anxiety on complex or difficult learning tasks in which competing error tendencies are stronger than the tendency to select the correct choice. An example of a difficult item having strong competing tendencies is one where, initially, the student's tendency to select an incorrect choice is stronger than his tendency to select the correct choice, or one where, initially, the student's tendency to select the correct choice was about equal to his tendency to select one or more incorrect alternatives. Based on this theory, the greater the anxiety at the time of feedback, the greater the strength of all the student's response tendencies and the greater the competition among the correct and incorrect alternatives. Thus, if the student has more anxiety at the time of immediate feedback than he does at the time of delayed feedback, it follows that his retention after delayed feedback will be superior to that after immediate feedback.

According to the Trait-State Anxiety Theory (Spielberger, 1966, 1971), it is essential to distinguish between "state" anxiety and "trait" anxiety. State anxiety refers to a complex response condition that varies in intensity and fluctuates over time; it is characterized by feelings of tension and apprehension and by activation of the autonomic nervous system. Trait anxiety refers to relatively stable individual differences in anxiety proneness.

Several recent studies of computer-assisted instruction have examined state anxiety and have supported the contention that periodic state anxiety measures can be used to investigate the relationship between anxiety and performance (Leherissey, O'Neil, & Hansen, 1971; O'Neil, 1972; O'Neil, Speilberger, & Hansen, 1969; Leherissey, O'Neil, Heinrich, & Hansen, 1973). These studies used the 20-item State Anxiety Scale from the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) to measure state anxiety of students when learning materials presented via computer-assisted instruction. Results showed that higher levels of state anxiety were associated with the more difficult learning materials, and that subjects with high state anxiety made more errors in these materials. These findings indicate that the State Anxiety Scale can be used to investigate the hypothesis that the delay retention effect is related to different amounts of anxiety experienced at immediate and delayed feedback.

Purpose

The purpose of the present experiment was to determine whether the findings of the studies described above were applicable to computer-managed testing in a college course. The specific purpose was to determine whether immediate or delayed feedback had differential effects on later retention.

METHOD

Experimental Conditions

The experimental conditions of this experiment were made as similar as possible to those in which the delay retention effect occurred:

1. Students were given a computer-managed test covering regular class material.

2. Both the computer-managed test and the retention (criterion) test were in a multiple-choice format.

3. Informative feedback was the re-representation of each item, with an indication of the correct alternative.

4. Four feedback conditions were provided: one immediate feedback (2-second delay), two delayed feedback (20-minute and 24-hour delay), and one no feedback (control). In previous experiments, feedback with a 20-minute delay was referred to as "immediate," since it was provided immediately after the student had responded to an entire set of items. However, since this procedure produces a considerable delay between response to an item and its feedback, it will be classed as delayed feedback in this report.

5. Twenty-four hours after taking the computer-managed test, students assigned to all feedback conditions were given the score obtained on the test (i.e., total number and percent correct).

6. Between 1 and 3 weeks later, students were tested for retention of this material.

7. For both the computer-managed test and the criterion test, two additional measures were used: state anxiety and confidence ratings. State anxiety was measured to determine whether there was a difference among the delayed feedback conditions as to the amount of anxiety experienced by the student when feedback was presented. Estimates of a subject's confidence in his response were used to provide a more continuous measure of feedback effectiveness than that provided by percent correct on the tests.

Subjects and Design

The 112 students in four sections of the upper-division course in Child Psychology at the California State University at Chico participated in the experiment. The course was taught by two instructors (A and B), each responsible for two sections. The students assigned to each instructor were assigned to a randomized block design; the blocking variable was the total score on the first two tests in the course.[1]

[1]The correlation between the blocking variable and the score on the computer-managed test was $r = .51$ for each instructor; the correlation between the blocking variable and the multiple-choice items that were in both the computer-managed and the criterion tests was $r = .36$ for each instructor. The degree of these relationships justifies the use of a randomized blocks design (see Hayes, 1963, p. 455).

Within the blocks, students were randomly assigned to one of the following feedback conditions:

1. No feedback provided (control group).

2. Feedback item-by-item, about 2-seconds after the student makes each response.

3. Feedback with about a 20-minute delay interval (provided after the student has responded to an entire set of items).

4. Feedback provided after a 24-hour delay interval.

The number of students assigned to each feedback condition is shown in Table 1.

Table 1

Students Assigned to Various Feedback Intervals

Instructor	Block	None (Control)	Feedback Intervals 2-Second	20-Minute	24-Hour	Total
A	I	4	4	3	3	14
	II	3	3	3	4	13
	III	4	4	3	3	14
	IV	3	3	4	3	13
Total--Instructor A		14	14	13	13	54
B	I	4	4	4	3	15
	II	2	4	3	5	14
	III	4	2	4	4	15
	IV	2	4	3	5	14
Total--Instructor B		13	14	14	17	58
GRAND TOTAL		27	28	27	30	112

Both Instructors A and B used Biehler (1976) as the text, and both required participation in the computer-managed and criterion tests as part of the course. However, the way the two instructors conducted their classes differed in several respects which may be relevant to the results:

1. Instructor A did not include grades obtained in the two tests in the course grade; Instructor B counted the highest score obtained in the two tests as part of the final grade.

2. Instructor A gave five exams throughout the semester, discarding the one with the lowest grade. Exams were on the book and lectures, they were not discussed in class, and there were no make-up exams until the final week.

3. Instructor B gave seven exams throughout the semester, on the text only. After completing each exam, the student took it to the instructor to be graded and then checked the graded exam with a feedback book, which provided an explanation for each answer. If desired, the student could take an alternate make-up of an exam on any exam day, once each week. The final course grade was the average of the highest scores obtained on six exams.

Testing Materials

Computer-Managed Test

The computer-managed test consisted of 30 multiple-choice (four alternatives) items (Form A, Quiz 6--Biehler, 1976). Using Kuder-Richardson Formula 20, the reliability of this test was $\underline{r} = .75$. Informative feedback consisted of the re-representation of the item, including the stem and all four alternatives, with a statement indicating the correct alternative (e.g., "Alternative C is correct").

This test was administered using an interactive test program of the SOCRATES system on a Digital Equipment Corporation model PDP 11/45. The system, along with four Teleray Corporation cathode ray tube (CRT) terminals and a shared hard-copy printer, was housed in the Education Psychology Building.

Criterion Test

The criterion test consisted of two parts. The first part consisted of 47 multiple-choice (four alternatives) items, including the 30 that were in the computer-managed test and 17 additional items on the same material (Form B, Quiz 6--Biehler, 1976). The items were presented in a random order in a dittoed test format. For the 30 items that were in both tests, the alternatives were in the same order on both tests and on the feedback presentation.

The second part consisted of 10 questions requiring a word or short phrase as an answer. Five of these questions were adapted from selected items in the computer-managed test; that is, they consisted of the stem plus the correct alternative, but a phrase or word was missing. Five were different items over the same material.

Additional Measures

For both tests, two additional measures were used--state anxiety and confidence ratings. These measures are described in the following paragraphs.

State Anxiety Measures

The student's anxiety was measured by the short form of the 20-item State Anxiety Scale of the State-Trait Anxiety Inventory (Spielberger et al., 1970). This form consists of those five items having the highest correlations with the remaining 15 items of the scale. Instructions for completing this scale were either *standard* (i.e., "Indicate how you feel right now") or retrospective (i.e., "Indicate how you felt during the task you have just finished"). As indicated previously, the purpose of this scale was to determine whether there was a difference among the feedback conditions as to the amount of anxiety the student was experiencing at the presentation of feedback.

Confidence Ratings

Asking subjects to indicate their degree of confidence in each response made provides a more continuous measure of performance than correctness or errors, and may prove more sensitive to feedback manipulations. Bayes' Theorem of conditional probability states that the ratio of the conditional probabilities of two events, given some datum (the posterior odds ratio), is equal to the likelihood ratio of that datum under those events times the ratio of the unconditional prior probabilities of those two events:

$$\frac{p\ (E_1/D)}{p\ (E_2/D)} = \frac{p\ (D\ /\ E_1)}{p\ (D\ /\ E_2)} \times \frac{p\ (E_1)}{p\ (E_2)}$$

or:

$$\text{Odds Ratio}_{\text{posterior}} = \text{Likelihood Ratio} \times \text{Odds Ratio}_{\text{prior}}$$

To illustrate, if schizophrenics outnumber depressives 2:1 (prior odds ratio) and a particular test score is three times as likely for a schizophrenic than a depressive (likelihood ratio = 3:1), then a new individual with this test score is very probably a schizophrenic (posterior odds ratio = 6:1).

Traditionally (Edwards, Lindman, & Savage, 1963), the logarithm of this expression is taken to assure the additivity of the various components:

$$LOR_{\text{posterior}} = LLR + LOR_{\text{prior}}$$

The LLR is taken to represent the impact or potency of a particular datum; that is, a very potent datum produces a greater change in probability than a less potent datum.

For the present experiment, it appeared that a subjective probability of the "correctness" of an answer could be obtained immediately prior to any feedback (a prior probability of being correct), as well as during the recall situation following feedback (a posterior probability of being correct). Use of Bayes' Theorem as outlined above, then, should result in a measure of the impact or potency of the particular feedback interval (a likelihood ratio) since:

$$LLR = LOR_{posterior} - LOR_{prior}$$

Two aspects of this formulation are important. First, such a procedure should produce a continuous measure of feedback effectiveness rather than the more traditional dichotomous right-wrong. Second, since the LLR is a "difference" score, individual differences between students in response level should not affect this ratio.

To obtain this measure, for the multiple-choice items in both tests, the student was requested to indicate his degree of confidence as to the correctness of his response. Instructions for completing the confidence ratings presented a 3-inch scale, with one end-point labelled 1, meaning "Guess"; and the other, 9, meaning "Certain." For each multiple-choice item, the student would first choose an alternative and indicate his confidence in that choice by typing or writing a number from 1 to 9.

In a four-alternative situation, the probability of a correct random response was .25; thus, a confidence level of 1 (Guess) for either a correct or an incorrect response was set at $p = .25$. Since it was assumed that the remaining values, up to a value of 9, were linear with p, correct responses were assessed at equal intervals from .25 to .99; and incorrect responses, at equal intervals from .25 to .01. For both tests, the $\log_{10}$ of the ratio p/q was taken as the response measure. For example, a 1 confidence estimate was associated with $p = .25$ and $q = .75$; the odds ratio was .25/.75 = .3333; and the log odds ratio was $\log_{10} .3333 = 0.47712$. As noted above, this value for the computer-managed test is an LOR_{prior} and that for the criterion test is an $LOR_{posterior}$.

Procedure

Computer-Managed Test

Each student reported to the computer testing room to take the computer-managed test at a time scheduled by him sometime between the 11th and 13th weeks of the semester. At this session, the student was: (1) given the short form of the A-State Scale with standard instructions (i.e., "Indicate how you feel right now"), (2) briefed on the use of the computer, (3) given two sample items, and (4) presented the test by the computer, one item at a time. The student responded to each item by typing the letter of his chosen alternative and the number of his confidence rating.

Students in the 2-second delay interval group (N = 28) received informative feedback immediately after they had responded to each item, with instructions to study the feedback but not to write anything. Students in the 20-minute delay interval group (N = 27) received feedback on all 30 items in a series after they had completed the test, with instructions to study the feedback but not to write anything. Students in the 24-hour delay interval group (N = 30) and in the control group (N = 27) received no feedback at the first session.

Students in all groups were given the short form of the A-State Scale immediately after they had completed the test. However, this time, the instructions were retrospective (i.e., "Indicate how you felt during the task you have just finished"). Students in the 20-minute delay interval group were given the scale again after they had received and studied their feedback.

Twenty-four hours after a student had taken the computer-managed test, he returned to the computer testing room. If he were in the control, 2-second delay interval, or 20-second delay interval group, he was given the short form of the A-State Scale with standard instructions, checked into the computer and given a report of his total score on the quiz (the number and the percentage correct), and then dismissed. However, if he were in the 24-hour delay interval group, he was:

1. Given the A-State Scale with standard instructions.

2. Checked into the computer and presented with informative feedback on the entire set of items he had completed 24 hours previously, with instructions to study the feedback but not to write anything.

3. Given the A-State Scale again, but this time with retrospective instructions.

4. Checked into the computer again to receive the total score obtained on the test.

5. Dismissed.

On the computer-managed test, students had complete control of the time that (1) each question was exposed both before and after they had recorded their response and confidence rating and (2) informative feedback for each item was exposed--up to 1 minute, when feedback was removed automatically. However, the testing session was programmed so that no student could proceed to the next item or to the next task until he had responded to the preceding ones. Also, he could not change an answer once it had been recorded and he could not review previous questions. A research assistant was present at all times during both computer sessions to check each student into the computer, to administer the A-State Scales, and to proctor the tests.

Criterion Test

The criterion test was given during regular class periods, at least 1 week and no longer than 3 weeks after the student had taken the computer-managed test. The part consisting of 10 short-answer items was given and collected before the part containing the 47 multiple-choice items was distributed. For each of these latter items, the student wrote the letter of his chosen alternative and the number of his confidence rating. After both tests were completed, the A-State Scale with retrospective instructions was administered. Finally, the student completed a questionnaire on the amount of studying he had done before and after the computer-managed test and his reactions to taking the test on the computer.

Analyses

Analyses were conducted to determine whether there were any differences between the feedback groups in (1) performance on multiple-choice items included in both tests, multiple-choice items included in the criterion test only, and short-answer "same" and "different" items; (2) time required to answer the items and to study informative feedback; and (3) anxiety experienced before and after testing. These analyses are described in detail in the following section.

RESULTS

Performance on Multiple-Choice Items Appearing in Both Computer-Managed and Criterion Tests

Student performance on the 30 multiple-choice items that appeared in both the computer-managed and criterion tests was of primary interest. Thus, for these items, analyses were conducted to determine the effect of the various feedback conditions on the following:

1. The mean number correct in both tests.

2. The proportion of items that were correct on the criterion test which had been either correct or incorrect on the initial test.

3. The degree of change in confidence ratings from the computer-managed test to the criterion test.

These analyses are described in the following paragraphs.

Mean Number Correct in Both Tests

Figure 1 shows the mean number of "same" multiple-choice items correct on both tests for all feedback conditions. Initially, separate unequal n randomized block analyses of variance (ANOVAs) of each test were conducted, indicating a significant effect of blocks on both tests (Computer test, $F\ (3,80) = 11.37$, $p < .01$ and criterion test, $F\ (3,80) = 4.57$, $p < .01$). However, since none of the interactions between blocks and either instructor or feedback conditions was significant, an overall unequal n ANOVA for these two test measures (computer and criterion test scores) was conducted with two between-group variables, instructor and feedback conditions. The overall effect of instructor, as well as the interactions among instructor, feedback, and test was not significant. However, there was a significant increase in the mean number correct from the computer-managed test to the criterion test ($F\ (1,104) = 48.56$, $p < .01$).

Analysis of the simple main effects of feedback conditions conducted at each test level showed no significant effect on the computer-managed test ($F\ (3,104) = .40$). However, the feedback conditions did have a significant effect on the criterion test ($F\ (3,104) = 6.70$, $p < .01$). When this effect was analyzed by three a priori planned orthogonal comparisons, the mean correct for the immediate feedback condition (2-second interval) and the two delayed feedback conditions (20-minute and 24-hour delay intervals) combined was significantly greater than that for the no feedback condition ($F\ (1,104) = 15.56$, $p < .01$). Also, the mean correct for the two delayed feedback conditions combined was significantly greater than that for the immediate feedback condition ($F\ (1,104) = 4.54$, $p < .05$). There was no significant difference between the 20-minute and 24-hour delayed feedback conditions.

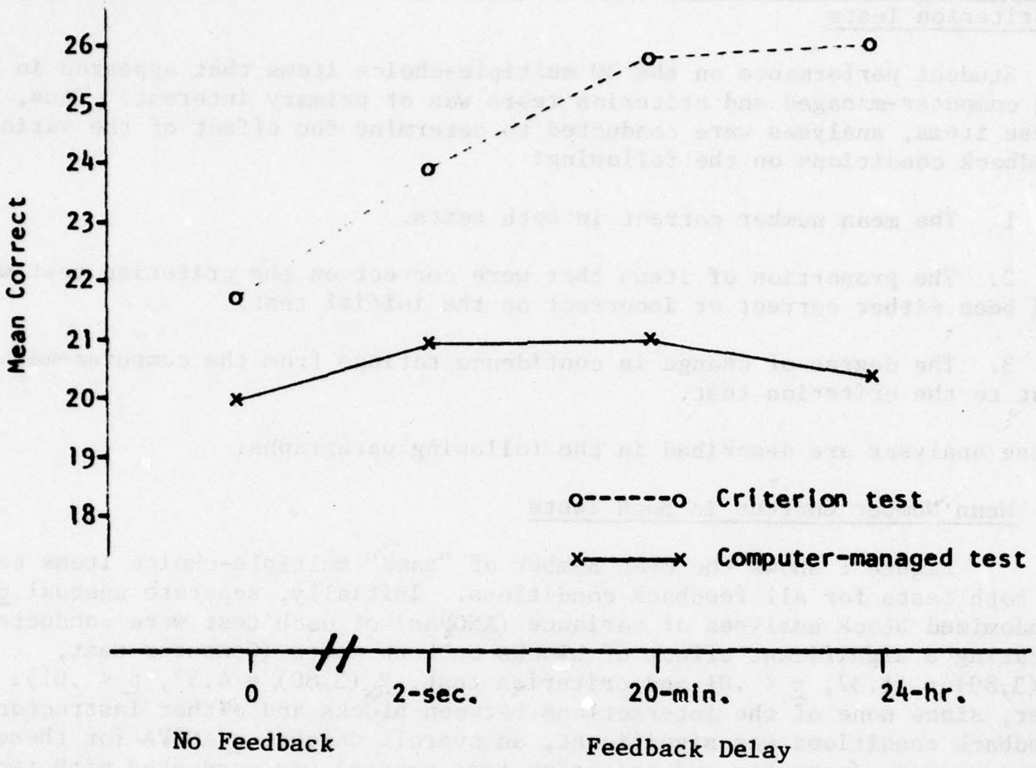

Figure 1. Mean number of "same" multiple-choice items correct on computer-managed and criterion tests for each feedback condition.

Proportion Correct on the Criterion Test That Had Been Correct or Incorrect on the Computer-Managed Test

The effect of feedback conditions on the proportion of items that were correct on the criterion test was analyzed separately for items that were correct and that were incorrect on the computer-managed test. Each student's responses on the computer-managed test were divided into those that were correct or incorrect, and the proportions of each of these that were correct on the criterion test were compared among the four feedback conditions.

For the proportion of items that were correct on both tests, there was no significant difference among any of the four feedback conditions (no feedback = .89; 2-second, 20-minute, and 24-hour delay intervals = .90, .93, and .93 respectively). This indicates that the performance of the control group, which received no informative feedback, did not differ from that of the groups receiving feedback on this measure. However, for the proportion of items that were wrong on the first test and right on the second test, effects of feedback conditions were similar to those obtained in the analysis of mean number of items correct on the criterion test. That is, the proportion for the 24-hour delayed feedback condition (.72) did not differ from that for the 20-minute delayed feedback condition (.67), $\underline{z} = 1.24$; the proportion for the two delayed feedback conditions was significantly higher than that for the 2-second immediate feedback condition (.53), $\underline{z} = 3.12$, $\underline{p} < .01$; and the proportion for the immediate feedback condition was significantly higher than that for the no feedback condition (.39), $\underline{z} = 3.34$, $\underline{p} < .01$.

Change in Confidence Ratings from Computer-Managed to Criterion Test

As noted previously, the transformation of the confidence ratings given in the initial computer-managed test yielded a log prior odds ratio (LOR_{prior}); and those given in the criterion test, a log posterior odds ratio ($LOR_{posterior}$). In accordance with Bayes' Theorem, the difference between these two values is the log likelihood ratio (LLR), which should assess the impact or potency of the feedback condition per se. Consequently, LLRs were established for each criterion--the initial response set for each item (i.e., wrong-wrong, wrong-right, right-wrong, and right-right). These LLRs were then analyzed by an unweighted means ANOVA involving the four feedback conditions, crossed with the four response classifications. It is important to note that, since a difference score was generated, it could be assumed that between-subject differences had been eliminated and that only treatment and within-subject error affected the ratio. Even if this assumption is not entirely valid, the results cited here are conservative and should not lead to excessive Type 1 errors.

The LLR for each of the four feedback conditions and response classifications is presented in Table 2. Although there was a significant effect of response classification ($\underline{F}$ (3,3344) = 1342.23, $\underline{p}$ < .01), this effect is meaningless since it merely reflects the differences between the types of responses per se; for example, wrong-right responses are positive by definition, and right-wrong responses, negative by definition. There was a significant effect of feedback conditions ($\underline{F}$ 3,3344) = 5.58, $\underline{p}$ < .01), as well as a significant interaction between feedback condition and response classification ($\underline{F}$ (3,3344) = 8.15, $\underline{p}$ < .01). When simple main effects of feedback conditions for each response classification were analyzed, there were no significant differences among feedback conditions for wrong-wrong responses ($\underline{F}$ (3,3344) = .84) or for right-right responses ($\underline{F}$ (3,3344) = .69). Thus, the confidence ratings of these two response classifications were not affected by either different feedback conditions or the presence or absence of feedback.

Table 2

Change in Confidence Ratings from Computer-Managed to Criterion Test (Log Likelihood Ratio)

Response Classification	Feedback Conditions			
	0	2-sec.	20-min.	24-hr.
Wrong-Wrong	-0.1971	-0.0298	-0.1438	-0.0616
Right-Right	0.3162	0.3384	0.3020	0.4542
Wrong-Right	1.8153	1.6250	2.0500	2.4181
Right-Wrong	-1.5762	-1.4716	-2.1199	-1.7351

Feedback conditions did have a significant effect for wrong-right responses ($\underline{F}$ (3,3344) = 16.76, $\underline{p}$ < .01) and for right-wrong responses ($\underline{F}$ (3,3344) = 11.74, $\underline{p}$ < .01). Thus, for both types of responses, a Newman-Keuls analysis was conducted on the differences among the means. For the wrong-right responses, there was a significantly greater increase (at the .01 level) in confidence ratings for the 24-hour feedback group than for each of the other three groups, and the increase for the 20-minute feedback group was significantly greater than that for the 2-second group. The other comparisons were not significant. Since a wrong-right response is one where an initially wrong response is changed to a correct one, an increased change in confidence rating is desirable. Thus, in this study, providing feedback after a 24-hour delay was significantly superior to all other feedback conditions. Even providing feedback after a 20-minute delay is superior to providing immediate (2-second interval) feedback.

Results of the Newman-Keuls analysis of the right-wrong responses showed that the 20-minute feedback group had a significantly greater change (at the .01 level) in confidence ratings than each of the other three groups. Since a right-wrong response is one where an initially right response is changed to a wrong one, an increase in the confidence rating in this case is not desirable. Thus, this analysis indicates that providing feedback after a 20-minute delay is significantly worse in this respect than any other feedback condition. It is particularly interesting to note that no differences on this measure were found between the 24-hour delayed feedback group and the no feedback (control) group.

In addition, an analysis of the overall level of change in confidence ratings was made for both wrong-wrong and right-right responses. For wrong-wrong responses, the mean change in confidence ratings appears to be small (F (3,3344) = 3.50, $p < .05$); however, it is significant statistically. This indicates that, even though the response was wrong in both tests, the student's confidence that his answer was, in fact, correct was significantly reduced; and this reduction occurred whether or not feedback of the correct answer was provided.

For right-right responses, there was a highly significant increase in the overall level of change in confidence ratings. Since this increase occurred for the no feedback (control) group, as well as for those receiving feedback, the time interval seems to be primarily responsible.

Item Difficulty and Feedback

Further, an analysis was conducted on the effect of feedback conditions on the mean correct on the criterion test as a function of the initial difficulty of the items. The items were divided into three sets--of 10 items each--based on the percentage of students who had each item correct on the initial computer-managed test. The mean correct on the criterion test for each feedback condition for each of three levels of difficulty is shown in Figure 2. An unequal n ANOVA on these data was conducted with the four feedback conditions crossed with the three levels of item difficulty, and both main effects were significant: feedback, F (3,108) = 7.69, $p < .01$; and item difficulty, F (2,216) = 58.98, $p < .01$. An analysis of the simple main effects of feedback conditions at each level of item difficulty indicated that they had a significant effect only on the most difficult items (F (3,108) = 11.13, $p < .01$).

Additional *a priori* planned orthogonal comparisons indicated the same relationships among the feedback conditions for the most difficult items as found above: The performance of groups receiving feedback was superior to that of the no feedback (control) group (F (1,108) = 19.23, $p < .01$); the performance of the two delayed feedback groups (20-minute and 24-hour delay intervals) was superior to that of the immediate (2-second delay) feedback group; and there was no significant difference in the performance of the two delayed feedback groups. This analysis is at least partly redundant with the analysis indicating that the effects of feedback conditions occurred only on wrong-right responses, since the most difficult items selected were those that were most often wrong on the computer-managed test.

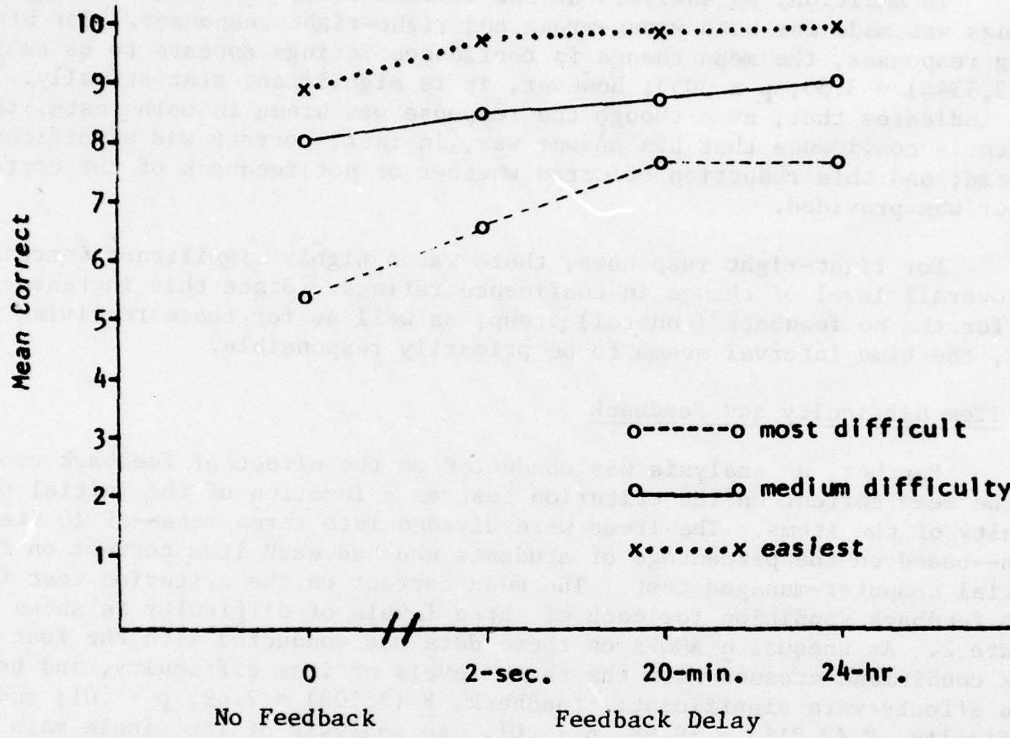

Figure 2. Mean correct on criterion tests for three levels of item difficulty for each feedback condition.

Performance on Criterion Test Items

"Different" Multiple-Choice Items

As indicated previously, the criterion test included 17 multiple-choice items that were not in the initial computer-managed test. An ANOVA was performed on the mean number correct of these "different" items, using three between-group variables: instructor, blocks (I through IV for each instructor (see Table 1)), and feedback conditions. Results indicated that there was a significant difference between the two instructors (F (1,80) = 4.09, p < .05), as well as a significant effect of blocks (F (3,80) = 2.96, p < .05). Further analysis indicated that the students in blocks I and II had significantly higher scores than those in blocks III and IV (F (1,80) = 7.51, p < .01). The effect of feedback conditions was not significant (F (3,80) = .33).

Short-Answer Items

The criterion test also included 10 short-answer items, five that were adapted from multiple-choice items in the computer-managed test, and five others on the same material. Responses to these items were scored on a scale from 0 to 4; thus, the student could have scores of 20 on the five "same" items and 20 on the five "different" items.

This part of the criterion test was scored as follows:

1. An initial judge made a detailed key for the test.

2. Two different judges independently scored all of the tests.

3. Discrepancies between the two scores on any item were resolved by the initial judge.

An unequal n ANOVA was conducted on these data, with three between-group variables: instructor, feedback conditions, and blocks. The scores for the "same" and "different" items were repeated measures. Results showed that the mean for the "same" items (12.65) was significantly higher than that for the "different" items (10.77), F (1,80) = 12.21, p < .01; and that blocks had a significant effect (F (3,80) = 8.69, p < .01). The effect of feedback conditions (F (3,80) = .97) and the other comparisons were not significant.

Amount of Time Spent on Computer-Managed Test

For each student, the length of time each item was exposed on the computer-managed test was recorded, and the total time spent on the 30 items was analyzed by an unequal n ANOVA with two between-groups variables, feedback conditions and blocks. There were no significant effects on this measure.

Also, for each student in the three groups receiving feedback, the time spent on feedback for each item was recorded. An unequal n ANOVA was conducted on the total time spent on feedback for the 30 items, with two between-groups variables--feedback conditions and blocks. There were no significant effects.

State Anxiety Measures

As indicated in the METHOD section, the short form of the State Anxiety Scale was administered at the following times:

1. To students in all feedback groups before they took the computer-managed test (with standard instructions), and immediately after they had completed the test (with retrospective instructions). (At this time, the immediate (2-second delay) feedback group had received and studied their feedback.)

2. To students in the 20-minute delayed feedback group after they had received and studied their feedback (with retrospective instructions).

3. To students in all feedback groups when they returned to the computer testing room 24 hours later to receive their score (with standard instructions).

4. To students in the 24-hour delayed feedback group after they had received and studied their feedback and before they received their test score (with retrospective instructions).

5. To students in all feedback groups after they had completed the criterion test (with retrospective instructions).

Each A-State questionnaire was scored by two independent judges. On the first measure, taken before the computer-managed test was administered, there was a significant difference among the feedback groups (F (3,80) = 4.65, $p < .01$). Although the three a priori orthogonal comparisons planned for analysis of the later measures were not directly meaningful to this first measure, they were conducted to provide a more accurate basis for interpreting the later comparisons. Results indicated that the A-State scores for the 2-second immediate feedback group were significantly higher than the combined scores for the 20-minute and 24-hour delayed feedback groups (F (1,80) = 12.62, $p < .01$). There was also a significant interaction between blocks and feedback conditions (F (9,80) = 2.12, $p < .05$).

Since there was an initial difference in the A-State measure among the feedback conditions, an analysis was made of the difference between the initial A-State measure and the measure taken after information feedback had been presented. For the 2-second immediate feedback and no feedback groups, the difference was computed between the initial measure and the measure taken directly after the computer-managed test had been completed (see 1 above). For the 20-minute and 24-hour delayed feedback groups, the difference was computed between the initial measure and the measure taken directly after they had received and studied their feedback (see 2 and 4 above). An unequal n ANOVA conducted on these measures indicated no significant effects.

The means for the A-State measure taken when students returned to the computer testing room 24 hours after they had taken the computer-managed exam (see 3 above) were compared to determine if there was any difference between the groups who had received feedback on the day they had taken the computer-managed test (the 2-second immediate feedback and the 20-minute delayed feedback groups) and those who had not (the no-feedback and the 24-hour delayed feedback groups). Results indicated that the mean A-State measure for those who had not received feedback (8.30) was significantly higher than that for those who had (7.24), F (1,79) = 4.41, $p < .05$.

Finally, an unequal n ANOVA was conducted on the A-State measures taken for all feedback groups after they had completed the criterion test. There were no significant effects.

Based on these findings, the results of the A-State measures must be considered as inconclusive.

Reactions Toward Computer-Managed Test

Responses to the question concerning student reactions toward the computer-managed test were categorized as positive (41%), negative (23%), or mixed or neutral (36%). A chi square analysis indicated that there were no significant differences among the feedback conditions as to relative response frequencies in these three categories.

CONCLUSIONS AND DISCUSSION

Results of this study have confirmed a consistent finding of previous laboratory experiments with computer-managed tests in an educational setting: Long-term retention of course material following immediate informative feedback is not superior to that with delayed informative feedback. As stated earlier, in previous experiments, the term "immediate feedback" was used for conditions where feedback is presented item-by-item immediately (approximately 2-second delay) after the student makes each response, and to conditions where it is presented after the student has responded to an entire set of items (20-minute delay). In the present experiment, as in previous experiments, retention following either the 2-second or 20-minute feedback conditions was not superior to that following a longer delay. It is of particular interest that the present results occurred in an educational setting with no control over study of material and in classes in which instructional practices differed in seemingly important ways.

Present results indicate that retention performance following 20-minute or 24-hour delays was superior to that following a 2-second delay. This was produced by the proportion of items wrong on the initial test and right on the criterion test (wrong-right responses). This finding is consistent with the perseveration-interference interpretation of the delay retention effect (Kulhavy & Anderson, 1972). However, the fact that there was no difference in wrong-right responses between the 20-minute and 24-hour delayed feedback conditions raises questions about the length of time response traces perseverate and indicates a need for clarification of this interpretation.

The present results also support the interpretation that differences at retention are due to conditions of presentation of feedback per se and not to indirect effects such as increased motivation, studying, etc. Feedback conditions had no effect on "different" multiple-choice items in the criterion test, for which no feedback was provided. Also, groups receiving feedback did better on the criterion test than the no feedback (control) group.

The delay retention effect, however, was not as marked in the present study as it has been in previous laboratory experiments. There was no difference in performance on the multiple-choice criterion test between the 20-minute and 24-hour delayed feedback conditions, and feedback intervals--or even the presence or absence of feedback--had no effect on the short-answer criterion test. Although it is not clear why the longer 24-hour delay did not result in superior retention, it is true that several conditions of the present experiment "pushed the limits" of those in previous experiments. The initial performance level prior to feedback was relatively high and the retention interval for many students was longer than that previously used. Also, the reliability of both the computer-managed and criterion tests was relatively low. The fact that the delay retention effect occurred for only the items with the highest level of difficulty does suggest that a greater effect might occur with more difficult test items. However, it seems likely that the present results, or lack of results, on this measure are due to a combination of factors.

Using the confidence ratings, which were included to provide a more continuous measure of retention than the more traditional measures, retention following the longer 24-hour delayed feedback was superior to that of all other feedback conditions. This finding is consistent with the interpretation that students learn more when they receive information about the correct answer after a longer delay interval (Sturges, 1969; 1972), and with interpretations that retention is a function of both the level or depth of processing and the spread of encoding (e.g., Craik & Tulving, 1975). Improved retention following greater depth of processing and/or spread of encoding has been reported for both intentional and incidental learning, and has occurred in situations similar to that of the present study (Craik & Tulving, 1975). Also, Seamon, Murray, and Barclay (1976) report that both confidence ratings and accuracy are higher in an incidental learning recognition test after engaging in a semantic-orienting task (with meaningful words) rather than a structural-orienting task.

It is hypothesized that, after a longer delay interval, students engage in a more thorough semantic analysis of information presented at feedback, which accounts for the increased confidence ratings on the retention test. However, the finding that feedback intervals had no effect on the time spent in answering the items and in studying the feedback is not consistent with this hypothesis unless it is assumed that retention following informative feedback is a function of how information is processed, and that different kinds of processing take about the same amount of time. Clearly, more direct study of this problem is necessary.

Previous studies have reported that the relative impact of immediate and delayed feedback on later retention varies with the relationships among the forms of the initial item, the feedback presentation, and the criterion test item (Sturges, 1969; 1972). Thus, the interpretations and implications of the findings of the present report are limited since all items, in both tests, were presented in the same form (i.e., with the same alternatives in the same order). Another limitation of the present results is the fact that there was only one test session with informative feedback. No evidence currently exists on the relative effectiveness of immediate and delayed feedback for repeated computer-managed tests throughout a course.

RECOMMENDATIONS

It is recommended that:

1. Further research be conducted to extend the findings of the present study by comparing the relative effects of immediate and delayed feedback under other experimental conditions; for example, using different forms of feedback presentation and/or criterion test items, and conducting repeated computer-managed tests with informative feedback throughout the course.

2. Procedures that increase student concentration on informative feedback and thus presumably influence the level or breadth of processing be developed and evaluated.

3. The appropriateness of requiring immediate feedback to be provided in curriculum development (NAVEDTRA 106A, 1975, Phase III, Table III.4, p. 9) be reconsidered since such conditions may not be optimum for learning.

REFERENCES

Biehler, R. F. Child development: An introduction. Boston: Houghton-Mifflin, 1976.

Brackbill, Y. The impairment of learning under immediate reinforcement. Journal of Experimental Child Psychology, 1964, 1, 199-207.

Chief of Naval Education and Training. Interservice procedures for instructional systems development (NAVEDTRA 106A). Pensacola: Author, August 1975.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

Edwards, W., Lindman, H., & Savage, L. J. Bayesian statistical inference for psychological research. Psychological Review, 1963, 70, 193-242.

English, R. A., & Kinzer, J. R. The effect of immediate and delayed feedback on the retention of subject matter. Psychology in the Schools, 1966, 3, 143-147.

Hayes, W. L. Statistics for psychologists. New York: Holt, Rinehart, & Winston, 1963.

Kulhavy, R. W. Feedback in programmed instruction and text material (NPRDC Tech. Note 77-1). San Diego: Navy Personnel Research and Development Center, October 1976. (AD-A033 300)

Kulhavy, R. W., & Anderson, R. C. The delay-retention effect with multiple-choice tests. Journal of Educational Psychology, 1972, 63, 505-512.

Leherissey, B. L., O'Neil, H. F., Jr., & Hansen, D. N. Effects of memory support on state anxiety and performance in computer-assisted learning. Journal of Educational Psychology, 1971, 62, 413-420.

Leherissey, B. L., O'Neil, H. F., Jr., Heinrich, D. L., & Hansen, D. N. Effect of anxiety, response mode, subject matter familiarity, and program length on achievement in computer-assisted learning. Journal of Educational Psychology, 1973, 64, 310-324.

More, A. J. Delay of feedback and the acquisition and retention of verbal materials in the classroom. Journal of Educational Psychology, 1969, 60, 339-342.

Newman, M. I., Williams, R. G., & Hiller, J. H. Delay of information feedback in an applied setting: Effects on initially learned and unlearned items. Journal of Experimental Education, 1974, 42, 55-59.

O'Neil, H. F., Jr. Effects of stress on state anxiety and performance in computer-assisted learning. Journal of Educational Psychology, 1972, 65, 473-481.

O'Neil, H. F., Jr., Spielberger, C. D., & Hansen, D. N. The effects of state-anxiety and task difficulty on computer-assisted learning. Journal of Educational Psychology, 1969, 60, 343-350.

Phye, G., & Baller, W. Verbal retention as a function of the informativeness and delay of information feedback: A replication. Journal of Educational Psychology, 1970, 61, 380-381.

Sassenrath, J. M., & Yonge, G. D. Delayed information feedback, feedback cues, retention set, and delayed retention. Journal of Educational Psychology, 1968, 59, 69-73.

Seamon, J. G., Murray, P., & Barclay, C. Levels of analysis in recall and recognition memory. Proceedings of the meeting of the American Psychological Association, Washington, D. C., September 1976.

Spence, K. W. A theory of emotional based drive (D) and its relationship to performance in simple learning situation. American Psychologist, 1958, 13, 131-141.

Spielberger, C. D. Theory and research on anxiety. In C. D. Spielberger (Ed.), Anxiety and behavior. New York: Academic Press, 1966.

Spielberger, C. D. Anxiety as an emotion state. In C. D. Spielberger (Ed.), Anxiety: Current trends in theory and research. New York: Academic Press, 1971.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. Manual for the State-Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press, 1970.

Sturges, P. T. Verbal retention as a function of the informativeness and delay of informative feedback. Journal of Educational Psychology, 1969, 60, 174-177.

Sturges, P. T. Information delay and retention: Effect of information in feedback and tests. Journal of Educational Psychology, 1972, 63, 32-43.

Sturges, P. T., & Crawford, J. J. Retention with delayed informative feedback as a function of meaningfulness of material and form of feedback (Res. Rep. No. 05-02). Olympia, WA: State Department of Public Instruction, April 1964.

Surber, J. R., & Anderson, R. C. Delay-retention effect in natural classroom settings. Journal of Educational Psychology, 1975, 67, 170-173.

Taylor, J. A. Drive theory and manifest anxiety. Psychological Bulletin, 1956, 53, 303-320.

DISTRIBUTION LIST

Assistant Secretary of the Navy (Manpower, Reserve Affairs and Logistics)
Principal Deputy Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Chief of Naval Operations (OP-01CR), (OP-987H), (OP-991B), (OP-964D)
Chief of Naval Personnel (Pers-10c), (Pers-2B)
Chief of Naval Material (NMAT 08T244)
Chief of Naval Research (Code 450) (4)
Chief of Information (OI-2252)
Chief of Naval Education and Training (N-5), (N-2), (003), (00A)
Chief of Naval Technical Training (Code 015), (Code 016)
Chief of Naval Education and Training Support
Chief of Naval Education and Training Support (01A)
Director of Navy Laboratories
Commandant of the Marine Corps (Code MPI-20)
Commander Training Command, U.S. Atlantic Fleet (Code N3A)
Commander Training Command, U.S. Pacific Fleet
Commander, Naval Data Automation Command
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 Anti-Submarine Warfare Training Center, Pacific
Commanding Officer, Naval Education and Training Program Development Center (2)
Commanding Officer, Naval Development and Training Center (Code 0120)
Commanding Officer, Fleet Training Center, San Diego
Commanding Officer, Naval Aerospace Medical Institute (Library Code 12) (2)
Commanding Officer, Naval Damage Control Training Center
Commanding Officer, Naval Training Equipment Center (Technical Library)
Commanding Officer, Naval Education and Training Support Center, Pacific (Code N1B)
Commanding Officer, Naval Health Sciences Education and Training Command (Code 2)
Commanding Officer, National Naval Dental Center (Library)
Officer in Charge, Naval Education and Training Information Systems Activity, Memphis Detachment
Director, Training Analysis and Evaluation Group (TAEG)
Director, Defense Activity for Non-Traditional Education Support
Master Chief Petty Officer of the Force, U.S. Atlantic Fleet
Master Chief Petty Officer of the Force, U.S. Pacific Fleet
Master Chief Petty Officer of the Force, Naval Material Command (NMAT 00C)
Master Chief Petty Officer of the Force, Naval Education and Training Command (Code 003)
Library Operations Section, Library of Congress
Occupational and Manpower Research Division, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Personnel Research Division, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Technical Library, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Technical Training Division, Air Force Human Resources Laboratory, Lowry 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
Advanced Systems Division, Air Force Human Resources Laboratory,
  Wright-Patterson Air Force Base
Program Manager, Life Science Directorate, AFOSR
Army Research Institute for the Behavioral and Social Sciences
Coast Guard Headquarters (G-P-1/62)
Secretary Treasurer, U.S. Naval Institute
Military Assistant for Training and Personnel Technology, Office of the
  Under Secretary of Defense for Research and Engineering
Director for Acquisition Planning, OASD(MRA&L)
Defense Documentation Center (12)