1995

How and When Should Information Be Restudied?

William L. Cull
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Psychology Commons

Recommended Citation

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License. Copyright © 1995 William L. Cull
LOYOLA UNIVERSITY CHICAGO

HOW AND WHEN SHOULD INFORMATION BE RESTUDIED?

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL OF
LOYOLA UNIVERSITY CHICAGO
IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
WILLIAM L. CULL

CHICAGO, ILLINOIS
JANUARY 1995
ACKNOWLEDGMENTS

I would like to thank the members of my dissertation committee, Dr. Eugene Zechmeister, Dr. Patricia Tenpenny, Dr. Lois Leidahl, and Dr. Emil Posavac, and the Graduate School for their help in the preparation of this paper. I would especially like to thank Dr. Zechmeister for serving as the director of this dissertation, for mentoring me throughout graduate school, and for always making himself available as both a teacher and a friend. Through his guidance and his example, I have come to appreciate what is involved in being a successful educator.

I would also like to thank my family and friends for their support, especially my parents. To those of you, who have agreed to serve as subjects in pilot experiments, who have read drafts of manuscripts, or who have been willing to discuss issues of learning and memory, thank you. Without your help this journey would have been much more difficult.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** ................................................................. iii

**LIST OF ILLUSTRATIONS** ......................................................... v

**LIST OF TABLES** ................................................................. vi

**Chapter**

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

- Study-Only Restudy ......................................................... 2
- Test-Only Restudy .......................................................... 6
- Test-Study Restudy ......................................................... 10
- Reprocessing Effects Integrated ......................................... 13

II. EXPERIMENT 1 ................................................................. 18

- Method ................................................................. 18
- Results and Discussion ............................................. 21

III. EXPERIMENT 1a ............................................................ 26

- Method ................................................................. 27
- Results and Discussion ............................................. 28

IV. EXPERIMENT 2 ................................................................. 37

- Method ................................................................. 37
- Results and Discussion ............................................. 41

V. EXPERIMENT 3 ................................................................. 48

- Method ................................................................. 48
- Results and Discussion ............................................. 52

VI. EXPERIMENT 3a .............................................................. 61

- Method ................................................................. 61
- Results and Discussion ............................................. 63

VII. GENERAL DISCUSSION ........................................................... 70

- How Should Information be Restudied? ................................. 72
- When Should Information be Restudied? ................................ 77

**REFERENCES** ................................................................. 86

**VITA** ................................................................. 95
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Experiment 2 Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials</td>
<td>45</td>
</tr>
<tr>
<td>2. Experiment 3 Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials</td>
<td>57</td>
</tr>
<tr>
<td>3. Experiment 3a Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials</td>
<td>67</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>Experiment 1 Recall Proportions as a Function of Type and Spacing of Repetitions</td>
</tr>
<tr>
<td>2.</td>
<td>Experiment 1a Recall Proportions as a Function of Type and Spacing of Repetitions</td>
</tr>
<tr>
<td>3.</td>
<td>Experiment 1a Recall Proportions as a Function of Type and Spacing of Repetitions Across Test Trials</td>
</tr>
<tr>
<td>4.</td>
<td>Experiment 2 Recall Proportions as a Function of Type and Spacing of Immediate Reviews</td>
</tr>
<tr>
<td>5.</td>
<td>Experiment 2 Recall Proportions as a Function of Type and Spacing of Immediate Reviews Across Test Trials</td>
</tr>
<tr>
<td>6.</td>
<td>Experiment 3 Recall Proportions as a Function of Type and Spacing of Delayed Reviews</td>
</tr>
<tr>
<td>7.</td>
<td>Experiment 3 Recall Proportions as a Function of Type and Spacing of Delayed Reviews Across Test Trials</td>
</tr>
<tr>
<td>8.</td>
<td>Experiment 3a Recall Proportions as a Function of Type and Spacing of Delayed Reviews</td>
</tr>
<tr>
<td>9.</td>
<td>Experiment 3a Recall Proportions as a Function of Type and Spacing of Delayed Reviews Across Test Trials</td>
</tr>
<tr>
<td>10.</td>
<td>A Comparison of Reprocessed and Nonreprocessed Items Across Experiments</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Finding optimal methods of acquiring and retaining information is an important goal within the fields of cognitive psychology and education. Many different learning methods have been investigated by researchers, and often these learning methods have been adopted by educators. Some of the optimal learning topics that have been explored include the investigation of imagery mnemonics (Lorayne & Lucas, 1974), the examination of the role of learning contexts (Smith, 1988), the exploration of the role of metacognition (Nelson & Narens, 1990), and the analysis of instructional technology developments, such as multimedia computer instruction (Mayer & Anderson, 1992).

It has been suggested, however, that, perhaps, the most important ingredient for learning is restudy. In most learning situations, especially school-like situations, learners study to-be-learned information on more than one occasion to ensure that the information is retained long enough to fulfill the learners' goals. Restudy efforts can be classified into two general categories: repetition and review. Repetition occurs when information is presented more than once within the same learning session. Review, on the other hand, occurs when information is re-presented
following the initial learning session.

Two important questions that need to be answered for both repetition and review of information are (a) How should information be presented to make restudy most effective? and (b) When should information be presented to make restudy most effective? These questions will be examined for both repetition and review in the present study. The question of how will be addressed in the present study by comparing three different methods of restudy that have been shown to be effective for learning: study-only restudy, test-only restudy, and test-study restudy. Research examining each of these types of restudy will be reviewed in turn. The question of when will be addressed in the present study by contrasting different spacing schedules of restudy. Particular attention will be paid to possible differences between expanding and uniform distributions of restudy, because of the high amount of publicity given to expanding spacings (see Baddeley, 1989; Banaji & Crowder, 1992).

Study-Only Restudy

Repeating information, by simply re-presenting information within a study session, or reviewing information, by re-presenting information within a separate study session, are the most typical methods of restudy. It has been demonstrated that when information is presented twice within a particular study session it is retained better than when it is only presented once (Reynolds & Glaser, 1964). This
finding has been labelled a repetition effect. It has also been shown that information that is presented twice across study sessions is better retained than information that is presented once (Ausubel, 1966). This finding has been labelled a review effect.

It is perhaps not surprising that repeating information during acquisition and reviewing previously learned information increase retention, since in both situations more total time is spent studying that information than nonrepeated or nonreviewed information (see Cooper & Pantle, 1967). However, review effects have been shown to have a greater effect on long-term retention than repetition effects when study time has been held constant (Reynolds & Glaser, 1964), suggesting that study time alone cannot sufficiently explain effective reprocessing.

The widely studied spacing effect demonstrates that the timing of re-presentation is also a critical factor for effective reprocessing. In the standard spacing effect, repeated presentations spaced throughout a learning session are more effective for learning than repeated consecutive presentations (Hintzman, 1974; Melton, 1970; Underwood, 1970). It has also been shown that multiple short presentations yield greater recall than a single longer presentation equal in duration to that of the shorter presentations combined (Cull, D'Anna, Hill, Zechmeister, & Hall, 1991). Thus, despite total time being held constant, spaced presen-
tations yield greater learning than massed presentations.

For an extended period of time, researchers investigated the spacing effect to understand the theoretical underpinnings of this effect (see Zechmeister & Nyberg, 1982). One interpretation that received a large amount of support was the attenuation of attention hypothesis. According to this interpretation, massed presentations are less effective than distributed presentations because the learner simply pays less attention to subsequent presentations when they occur shortly after an identical presentation. Research has suggested that learners may pay less attention because they falsely believe that the presented information has already been successfully learned and see no point to studying it further (Zechmeister & Shaughnessy, 1980). It is also possible that the presence of information in short-term memory minimizes the processing potential for a subsequent presentation, since no information needs to be searched for or retrieved from long-term memory (Jacoby, 1978); the generation of information from long-term memory is known to facilitate retention and later retrieval of information (Slamecka & Graf, 1978).

Spacing reviews of information has also been shown to be more effective than massing reviews immediately after initial study, at least for certain retention intervals. If the retention interval is short (e.g., a few minutes), "cramming" the study of information is roughly as effective
as distributing study (Underwood & Schultz, 1961); if the retention interval is long (e.g., a day), "cramming" results in lesser performance than spaced review (Keppel, 1967).

For both repetitions and reviews, researchers have further investigated what spacings benefit learning most. For repetitions, it has been found that increasing the lag between multiple repetitions of an item increases the probability of recalling that item (Hintzman, 1974). Researchers have been less successful in determining when reviews of information are most beneficial. For instance, it has been shown that a single review occurring shortly after study (e.g., one day) or occurring after long durations (e.g., one week) are equally helpful (Ausubel, 1966; Gay, 1973; Reynolds & Glaser, 1964). Attempts to determine the optimal spacing of multiple reviews have also been unsuccessful (Gay, 1973), although some educators have suggested, based more on logical reasoning than on empirical support, that reviews occurring at increasing intervals should be most effective (Lyon, 1914; Sones & Stroud, 1940).

A number of researchers have stressed the practical significance of distributed repetition and spaced review for education (Bjork, 1979; Dempster, 1986). This argument has been supported by research demonstrating the utility of spaced review in academic settings. For example, Smith and Rothkopf (1984) demonstrated that an eight-hour college statistics mini-course was more effective when the course
lessons were distributed rather than massed.

**Test-Only Restudy**

Educators have used tests of students' knowledge, ranging from questions asked during lectures to quizzes and examinations, to fulfill several educational purposes. Primarily, tests of students' knowledge assess how much students have learned and motivate students to study. In addition to these functions, tests provide students with opportunities to reprocess the originally studied material, assuming that the information can be successfully retrieved from memory. Test or retrieval opportunities that do not provide the learner with feedback concerning the correct response to those tests have been described as unreinforced test-trials or test-only trials.

As was the case for study-only reprocessing, information that is re-presented through test-only presentations is more likely to be remembered than information that is not re-presented. This has been demonstrated when test-only presentations were administered during the initial learning of information (Bartlett, 1977; Bartlett & Tulving, 1974; Madigan & McCabe, 1971; Modigliani, 1976; Young, 1971), when they were presented immediately after initial learning (Darley & Murdock, 1971; McDaniel, Kowitz, & Dunay, 1989; Runquist, 1983), and when they were presented more than a day after initial learning (Runquist, 1986a; Runquist, 1986b; Spitzer, 1939). These effects extend repetition and
review effects to learning situations using tests, rather than study-only presentations, for reprocessing. Although these results also conform well with the total-time interpretation of reprocessing, it is not at all clear whether study re-presentations and test re-presentations facilitate learning in the same way.

Research that has been conducted on test-only reprocessing has attempted to determine whether test repetitions simply provide additional study opportunities (if the information is remembered) or whether test repetitions provide more than additional study, such as retrieval practice or greater consolidation of information in memory. To answer this question, test-only opportunities have been directly compared to study-only opportunities. Some of the studies have found no difference between test-only and study-only presentations (Donaldson, 1971; Landauer & Eldridge, 1967; Whitten & Bjork, 1977); whereas, others have found higher recall for study-only presentations than for test-only presentations (Landauer, 1969; McDaniel & Masson, 1985). Although the idea that retrieval provides more than additional study is not supported by these results, it has been argued that the effects of retrieval are obscured by a sizable amount of information not being successfully retrieved for test-only repetitions (Wenger, Thompson, & Bartling, 1980).

Other studies have shown greater final recall for
information receiving test-only presentations rather than study-only presentations (Allen, Mahler, & Estes, 1969; LaPorte & Voss, 1975; Landauer & Bjork, 1978; Hogan & Kintsch, 1971; Wenger, Thompson, & Bartling, 1980). These results suggest that, at least in certain situations, retrieval benefits learning by providing more than additional study opportunities.

Thus, study-only and test-only presentations appear to have their own advantages. Test-only presentations presumably benefit the learning of information in ways that study-only presentations cannot, for example, by requiring greater processing of the learner (Schmidt & Bjork, 1992), by consolidating the memory trace (Landauer, 1969), or by providing practice at finding successful retrieval routes (Bjork, 1988). For test-only presentations to be effective, however, it is critical that the tested information is successfully retrieved from memory; whereas, study-only presentations can be effective regardless of whether the learner remembers seeing the information previously. The higher recall for test-only in comparison to study-only represents a retrieval effect (Bjork 1975). A retrieval effect as defined by this more stringent comparison, has yet to be demonstrated for review presentations.

Researchers have also varied the timing of single and multiple test-only repetitions in an attempt to find the most beneficial placement of repetitions. Whitten and Bjork
(1978) demonstrated that across short intervals (fifteen seconds) test-only repetitions were more beneficial when the lag between the initial study presentation and the re-presentation was longer. Landauer (1969) demonstrated that this lag effect occurs (despite the lesser probability that an item will be successfully remembered on the test-only repetition) because of the increased probability that a remembered item will not be forgotten between the repetition and final recall.

Landauer and Bjork (1978) attempted to determine the optimal schedule for providing a series of test-only trials. It was found that a series of test-only repetitions presented with expanding intervals between presentations was more effective than contracting, uniformly spaced, and massed presentation series. Intervals between repetitions were determined by the presentation of intervening items. For instance, one expanding series had gaps of one, four, and ten intervening items; whereas, a comparable uniform series had gaps of five, five, and five intervening items. Expanding test trials were thought to be optimal for learning because they shaped the learning process by decreasing the probability that the learner would forget information while still providing sufficiently hard tests of information. Testing information just as it is on the verge of being forgotten appears to provide the maximal processing of information (Bjork, 1988).
Recently, the expanding test series effect has been demonstrated in a variety of learning situations, ranging from learning names presented visually at a fixed rate to learning facts in a self-paced learning situation (Shaughnessy, Zechmeister, & Cull, submitted for publication). Prior to any replications of the effect, however, the expanding test series effect received considerable attention as a highly applicable memory finding (Baddeley, 1990; Banaji & Crowder, 1989).

Other researchers have also investigated the optimal spacing of test-only reviews. When the lags are measured in days, shorter lags rather than longer lags have been found to increase the likelihood of later recall (Runquist, 1986a). This has been especially true when multiple review tests are given for the same information (Runquist, 1986b; Spitzer, 1939). The expanding test series effect has not been examined using review tests.

**Test-Study Restudy**

Although less research has focused on test-study reprocessing, some recent studies have compared test-study trials to other reprocessing methods. Carrier and Pashler (1992) examined whether two ten-second study periods provided after an initial study trial would be as effective for learning as two ten-second test-study periods (five-second test / five-second study). They reasoned that this comparison provided the purest test of the retrieval effect, since retrieval
opportunities are only available for one condition but both
conditions receive the same overall amount of information.
Subjects' performance was reliably higher in the test-study
condition, suggesting further that retrieval benefits learning in ways that study presentations alone do not.

Test-study repetitions have the same advantages over
study-only repetitions as do test-only repetitions, but the
effectiveness of test-study presentations is not dependent
on successful retrieval as is the effectiveness of test-only
presentations. Moreover, the combination of a test and
study component may benefit learning by potentiating the
subsequent study of information through testing (Izawa,
1970). It has been shown, for example, that test-study
trials improve subjects' sensitivity to item difficulty and
overall retention of information within self-paced learning
situations (Cull & Zechmeister, 1994).

Shaughnessy et al. (submitted for publication) also
employed a test-study learning procedure in two of their
experiments, but for a different reason than did Carrier and
Pashler. In addition to investigating the external validity
of the expanding test series effect, Shaughnessy et al. were
interested in assessing the applicability of the effect.
They argued that for the expanding test series to be appli-
cable for school-like learning situations, it must be demon-
strated in a situation where students are provided with
feedback concerning the correct answers following retrieval
attempts. When the expanding test series effect was investigated using test-study repetitions, uniformly distributed presentations were found to be as effective as expanding presentations; both spacings of test-study presentations were more effective than expanding test-only presentations.

In a series of longitudinal studies, Bahrick has used a test-study review paradigm, spacing review at different intervals, in order to determine the optimal spacing for the long-term or permastore retention of information (Bahrick, 1979; Bahrick & Phelps, 1986; Bahrick, Bahrick, Bahrick, & Bahrick, 1993). In this learning paradigm, a learner is first given the opportunity to study a set of items. These items are then tested within the initial study procedure to ensure that the items have been learned. Any item not recalled correctly is again presented for study and testing. This procedure continues until all of the items are recalled correctly on one test trial. Review learning sessions are administered in the same way except that these sessions begin with a test.

Bahrick (1979) found that across a thirty-day retention interval, performance was best when a single review was spaced thirty days from initial learning rather than one day or no days from initial study. The difference between no days and one day was considerably larger than the difference between one day and thirty days. Interestingly, when asked to recall the same information seven years later, the recall
differences between the one day and thirty days conditions remained unchanged (Bahrick & Phelps, 1986).

Bahrick, Bahrick, Bahrick, and Bahrick (1993) learned foreign vocabulary using this same review procedure across a nine-year period. Items were reviewed many times throughout this period (more than twenty times) with the review sessions spaced fourteen, twenty-eight, or fifty-six days apart. Although it took more time on average to reacquire the words after the fifty-six-day reviews, these items were recalled better after retention intervals ranging from one to five years. Thus, it would seem that test-study reprocessing is beneficial for learning as a review method, even when the reviews are spaced across lengthy intervals.

Reprocessing Effects Integrated

A theoretical understanding of how restudy serves to facilitate learning and memory clearly is important to the general understanding of memory processes. However, because of the potential beneficial effects on memory, the applied implications of this research cannot be overlooked. By directly contrasting various methods of restudy that previously have been shown to be effective, a greater understanding of reprocessing and more accurate restudy prescriptions hopefully will be provided.

Theory. It should be evident that there are many different reprocessing effects and that these effects critically depend on specific timing and methods of restudy.
Nevertheless, there are some general principles that underlie or, at least, tie certain subsets of reprocessing effects together. One promising interpretation suggests that an item is more effectively processed, actually "reprocessed," when the additional processing of information is made more difficult for the learner (Schmidt & Bjork, 1992). That is, it is suggested that what makes processing easy during acquisition does not necessarily benefit later retention. This interpretation provides a reasonable account for spacing effects, as well as retrieval and delayed retrieval effects.

If a second repetition is delayed, even for a brief interval of time, it becomes considerably less likely that a given piece of information remains in short-term or primary memory (Peterson & Peterson, 1959). This requires the learner to once again bring the information into consciousness in order to process the information. If the information cannot be retrieved it will be processed as if it is new. The recovery of information from long-term memory or the processing of information as if it is new both require greater processing than processing information already in short-term memory. Moreover, retrieving an item from memory rather than simply studying the item again, would also be expected to require greater effort. If the act of retrieval is made more difficult by increasing the interval between study and a test trial, it will be more difficult to re-
trieve the desired information correctly, which likely results in greater processing when the information is successfully retrieved.

Cuddy and Jacoby (1982) suggest that the greater processing associated with spacing and retrieval is similar to the greater processing that is required when a learner must re-solve a previously encountered problem as opposed to remembering the solution to the problem. In their experiments, the probability of remembering a solution was systematically reduced by lengthening the time between presentations or varying the type of intralist filler information. Regardless of what reduced the probability of remembering a solution, when it was more difficult to solve a problem, that information was recalled better on a later retention test.

The beneficial effects of retrieval and spaced presentations, however, can also be explained at the memory stages of retention and retrieval as opposed to at acquisition. For example, a "retrieval-practice" perspective suggests that the act of searching and retrieving information from memory provides the learner with practice at establishing successful retrieval routes. Delayed retrieval is especially effective because it mimics final recall and provides the learner with a more realistic practice attempt. If an item is recalled, then that retrieval route is reinforced. If retrieval is unsuccessful, then the learner may attempt to
encode the information in a different, hopefully more productive way, if given an additional study opportunity.

Clearly, it is very possible that distributed practice and retrieval are beneficial for more than one reason. Furthermore, theoretical interpretations of memory effects based on acquisition as compared to interpretations based on retention and retrieval are notoriously difficult to reconcile (Watkins, 1991). Therefore, no attempt was made to directly contrast acquisition and retrieval interpretations of reprocessing in the present study.

The present study was designed to provide information for the theoretical integration of reprocessing in two different ways. First, an attempt was made to fill in some of the remaining gaps in our knowledge of reprocessing effects by investigating reprocessing under conditions that have yet to be examined. For instance, retrieval effects based on a comparison with a study-only control group have not been investigated for review tests, and the expanding test series effect has not been examined for review test-only or review test-study trials. Second, the present study examined repetition, review, spacing, retrieval, and expanding testing effects within the same learning paradigm with the goal of illuminating the similarities and differences among these effects.

Application. Researchers have been interested for a long time in the practical applicability of several of the
reprocessing effects previously discussed. It has been argued that spacing repetitions (Dempster, 1986), spacing reviews (Dempster, 1986; Smith & Rothkopf, 1984), providing retrieval opportunities (Bjork, 1979), and expanding the intervals between multiple test opportunities (Bjork, 1988; Landauer & Bjork, 1972; Rea & Modigliani, 1984) all are effective learning techniques that could and should be utilized within educational settings.

In order for psychologists to provide the best advice concerning these similar, albeit different, methods of restudy, a direct comparison of these different learning methods needs to be made. Experiments 1 and 1a were specifically intended to examine what type of repetitions and what spacings of repetitions were optimal for learning. In Experiment 2 the type and spacing of restudy were again compared, but for immediate reviews of information rather than for repetitions. Finally, Experiments 3 and 3a compared the type and spacing of reviews, but the reviews were spaced across days.
Although a great deal of research has focused on how and when to repeat information within a study session, there are many issues that still remain unresolved. Specifically, the external validity and applicability of both test-study retrieval effects and the expanding test series effect need to be tested further. Thus, the three types of restudy previously introduced, study-only, test-only, and test-study restudy, were compared within Experiment 1 using expanding, uniformly distributed, and massed spacings of repetitions. It was expected that the expanding spacing schedule would be more effective than the uniform spacing schedule for test-only repetitions (see Landauer & Bjork, 1978). It was also expected that the expanding spacing schedule would not be more effective than the uniform schedule for test-study or study-only repetitions and that test-study repetitions would be generally most effective (see Shaughnessy, Zechmeister, & Cull, submitted for publication).

Method

Subjects. A total of 57 subjects from the introductory psychology subject pool at Loyola University Chicago participated in the experiment.

Materials. All subjects were asked to learn, using an
IBM-compatible computer, a list of 32 paired-associate items of moderate difficulty according to Underwood's (1982) norms. An additional four items, also of moderate difficulty, were used as filler items. Only paired-associates that consisted of an uncommon word as the cue member and a common word as the response member were selected. This made the word pairs similar to those found in vocabulary study in that the to-be-learned word is uncommon, and the definition is made up of common words. Both words of each pair were five letters in length. An example pair is batik-lyric.

Procedure. Upon arrival, subjects were randomly assigned to study the items in one of three ways. In all three conditions, subjects were presented the same items for study on four different occasions; each presentation lasted a total of 8 s. In the first condition (study-only), the cue and response were presented for an 8-s period on each of the four presentations. In the second condition (test-only), both the cue and response members of the word pairs were presented for 8 s on the first presentation; but, on the following three presentations, the cue member alone was presented for the entire 8-s period. Finally, in the third condition (test-study), subjects were presented the cue and response for the full 8 s on the first presentation and the cue alone for 4 s followed by the cue and response for 4 s on the remaining three presentations. Subjects were encouraged to attempt to remember the appropriate response member
when presented the cue member alone in the test-only and test-study conditions.

In addition to the between-subjects variable, type of presentation, the spacing of presentations was also manipulated as a within-subjects variable. Subjects were presented all critical items according to their respective type of presentation (study-only, test-only, or test-study), but the items were randomly divided into four subsets (8 items). These subsets were randomly assigned to each of three spacing schedules (expanding, uniform, massed), leaving one subset to serve as control items that were presented one time. In the massed condition, all four presentations or tests of an item occurred consecutively; there were no intervening items. In the uniform condition, each repetition of an item occurred after a uniform gap of five other items. In the expanding condition, repetitions were also distributed but based on increasing intervals between items: a 1-item gap separated the first and second presentations of the item, a 5-item gap separated the second and third presentations, and a 9-item gap separated the third and fourth presentations. Both critical and filler items were used to create the gaps between item repetitions. Control items were only presented one time for 8 s with both cue and response presented. The items were assigned randomly to conditions for each subject, resulting in a different random order for each subject.
Once all the items had been studied, subjects were given a 1-min filler task (they were asked to read instructions for the retention tests) followed by a computerized cued-recall test of all the items. On the cued-recall test, the cue word from each of the originally studied items was presented along with a cursor box and a prompt that asked subjects to type in the appropriate response word. Subjects were given as long as they needed to enter a response on the final test, but they were not able to return to an item once a response had been entered or the item had been skipped. The order of the items again was random and thus different for each subject.

Results and Discussion

Across all experiments, an alpha-level of $p \leq .05$ was used for all statistical tests conducted. The proportions of correct responses on the final cued-recall test of Experiment 1 are listed in Table 1. These results were analyzed using a 3 X 4 mixed ANOVA. The independent variables, type and spacing of repetitions, were found to interact significantly with one another, $F(6,162) = 3.13, MSe = 1.50$. This interaction was explained through a series of simple effect analyses.

Although a significant main effect of repetition type was found, $F(2,54) = 3.47, MSe = 9.84$, simple effect analyses revealed that differences between test-study, test-only, and study-only repetitions were limited to expanding, $F(2,$
Table 1.—Experiment 1 recall proportions as a function of type and spacing of repetitions

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Type</th>
<th>Study-Only</th>
<th>Test-Only</th>
<th>Test-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanding</td>
<td>Type</td>
<td>.41</td>
<td>.44</td>
<td>.61</td>
</tr>
<tr>
<td>Uniform</td>
<td>Type</td>
<td>.40</td>
<td>.43</td>
<td>.67</td>
</tr>
<tr>
<td>Massed</td>
<td>Type</td>
<td>.20</td>
<td>.30</td>
<td>.28</td>
</tr>
<tr>
<td>Nonrepeated</td>
<td>Type</td>
<td>.10</td>
<td>.15</td>
<td>.20</td>
</tr>
</tbody>
</table>
For expanding items, test-only subjects did not recall reliably more items than study-only subjects did, test-study subjects recalled more items than study-only subjects did, $F(1, 54) = 4.82$, $MSE = 4.82$, and test-study subjects recalled slightly more items than test-only subjects did, $F(1, 54) = 3.69$, $MSE = 4.82$, $p = .06$. For uniform items, test-study subjects recalled significantly more items than study-only or test-only subjects, $F(1, 54) = 8.10$, $MSE = 4.45$, and no statistical difference was found between test-only and study-only subjects.

A significant main effect was also found for repetition spacing, $F(3, 162) = 72.53$, $MSE = 1.50$. For all three types of repetition, massed items were recalled more than non-repeated items [Study-only, $F(1, 18) = 8.47$, $MSE = .70$; Test-only, $F(1, 18) = 13.49$, $MSE = 1.03$; Test-Study, $F(1, 18) = 4.43$, $MSE = 1.00$], uniform items were recalled better than massed items [Study-only, $F(1, 18) = 17.37$, $MSE = 1.46$; Test-only, $F(1, 18) = 17.10$, $MSE = .56$; Test-Study, $F(1, 18) = 54.94$, $MSE = 1.99$], and expanding and uniform items were recalled equally well. As can be seen in Table 1, although the pattern of significance was the same for all repetition types, the difference between massed and distributed (uniform and expanding) items was greater for test-study repetition.

Experiment 1 was designed to determine what type and
what spacing of repetitions were most effective. The re-
sults of Experiment 1 revealed that, overall, repetitions
were important for learning and that test-study repetitions
were the most effective type of repetition examined. The
highest levels of recall resulted for test-study items in
the expanding (.61) and uniform (.67) conditions. Recall
was especially high in these conditions, because subjects
benefitted from both the positive effects of retrieval and
distributed practice. These findings provide a replication
of the test-study retrieval effect and suggest that distrib-
uted test-study repetitions are highly effective for learn-
ing.

The failure of expanding test-only repetitions to be
more effective than uniform test-only repetitions was unex-
pected, and the reason for this finding is not readily
apparent. One possibility is that the first tests for an
expanding schedule did not produce higher levels of recall
than the first tests for uniform items, and thus expanding
tests did not shape learning more than uniform tests. This
possibility was explored further in Experiment 1a.

For test-study repetitions, an expanding test series
effect was also not found. Expanding items, in fact, were
recalled less than uniform items although the difference was
not significant. This result is consistent with the results
of previous studies failing to find an expanding test series
effect for test-study items (Shaughnessy, Zechmeister, &
Cull, submitted for publication). Thus, the results of Experiment 1 do not support the claimed wide applicability of the expanding test series effect.
CHAPTER III

EXPERIMENT 1a

Because an expanding test series effect was expected but not found within Experiment 1, a second experiment was conducted to investigate again how the type and spacing of repetitions affect learning. The procedure used in Experiment 1 was changed in an attempt to increase the probability of obtaining an expanding test series effect. It was reasoned that expanding items may not have been more helpful than uniform items in Experiment 1 because subjects may not have recalled more expanding items than uniform items on the first recall attempt. That is, the quicker first test for expanding items may not have successfully increased performance on that test. Thus, in Experiment 1a, a greater number of total filler items were presented in order to increase the difficulty of recalling critical items. These filler items were presented fewer times than were the filler items in Experiment 1 so that the same spacing between critical items was maintained. Subjects also were asked to enter their responses using the keyboard so that performance on the repetition tests could be monitored. Finally, another variable, duration of initial study, was introduced; for one half of the items, the initial study presentation re-
mained 8 s in length, but for the other half, the initial study presentation was reduced to 4 s in length. This was also done in order to influence the probability of recalling information on the first retrieval attempt.

Method

Subjects. A total of 66 subjects from the same subject pool that was used for Experiment 1 participated in this experiment. No subject participated in both Experiments 1 and 1a.

Materials and procedure. The materials and procedures used in Experiment 1a were identical to those used in Experiment 1 with the following exceptions. First, 12 filler items (as opposed to 4) were used to provide spacings between critical items. Second, when items were re-presented in the test-only and test-study conditions, a cursor box appeared under the cue word, and subjects were asked to type in the appropriate response word. Subjects were given 8 s to enter a response in the test-only condition and 6 s to enter a response in the test-study condition. The cue and response were both presented for an additional 2 s for test-study items. Subjects were informed that it would be challenging to enter a response successfully during the initial study period, but they were assured that they would have unlimited time to enter their responses on the final recall test. All responses entered by a subject were saved for later analysis.
Finally, a third variable, duration of initial study, was manipulated in addition to the type and spacing of repetitions. Duration of study was a within-subjects variable. For each of the item sets that were randomly assigned to one of the four repetition spacings (expanding, uniform, massed, and single), half of the items were randomly chosen to have shorter first presentations. Short items were presented for 4 s on their first presentation; long items were presented for 8 s. Short and long items only differed for the first presentation; all subsequent presentations remained 8 s in length. The final cued-recall test was administered in the same way that it was in Experiment 1.

Results and Discussion

A 3 X 4 X 2 mixed ANOVA was used to examine the effects of type of repetition, spacing of repetition, and length of initial presentation on recall. Length of initial presentation was not found to have a significant effect on recall, or to interact with either of the other variables. Thus, the results, as summarized in Table 2, were combined for short and long initial presentations. A significant main effect was found for spacing, $F(3,189) = 68.11$, $MSe = .65$, and a marginally significant main effect was found for type of repetition, $F(2,63) = 2.57$, $MSe = 4.30$, $p = .09$. These effects were influenced, however, by a significant interaction between these variables, $F(6,189) = 2.66$, $MSe = .65$. Subjects in the test-study condition recalled reliably more
Table 2.--Experiment 1a Recall Proportions as a Function of Type and Spacing of Repetitions

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Study-Only</th>
<th>Test-Only</th>
<th>Test-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanding</td>
<td>.27</td>
<td>.38</td>
<td>.48</td>
</tr>
<tr>
<td>Uniform</td>
<td>.29</td>
<td>.34</td>
<td>.49</td>
</tr>
<tr>
<td>Massed</td>
<td>.19</td>
<td>.18</td>
<td>.19</td>
</tr>
<tr>
<td>Nonrepeated</td>
<td>.06</td>
<td>.09</td>
<td>.11</td>
</tr>
</tbody>
</table>
than subjects in the study-only condition did for expanding items, $F(1,63) = 7.03$, $MSe = 4.43$, and for uniform items, $F(1,63) = 5.56$, $MSe = 5.01$, but not for massed or non-repeated items. Subjects in the test-only condition, despite recalling more than study-only subjects did and less than test-study subjects did, were not significantly different than either group; the difference between the test-study and test-only conditions was, however, marginally significant for uniform items, $F(1,63) = 3.31$, $MSe = 5.01$, $p = .07$.

Spacing was again found to have the same pattern of significance for each type of repetition. Distributed items were recalled more than massed items [test-study, $F(1,21) = 43.73$, $MSe = 1.35$; test-only, $F(1,21) = 17.46$, $MSe = .95$; study-only, $F(1,21) = 6.79$, $MSe = 1.62$], and massed items were recalled more than single items [test-study, $F(1,21) = 4.34$, $MSe = 1.03$; test-only, $F(1,21) = 9.27$, $MSe = .63$; study-only items, $F(1,21) = 8.10$, $MSe = .55$]. The interaction of type and spacing of repetitions resulted from larger differences between distributed and massed items for test-study and test-only repetitions than for study-only repetitions. No significant differences were apparent between expanding and uniform items, although the proportions recalled did favor expanding items in the test-only condition.

In an attempt to explain further why test-study trials were more effective than test-only trials and why an expanding test series effect was not found for test-only repeti-
tions, an additional analysis was conducted that investigated recall performance for each repetition test, as well as for the final test. These results are summarized in Table 3. For test-only repetitions, spacing was found to interact with the trial of testing, $F(6,126) = 37.89$, $MSe = 1.32$. As was originally expected, expanding items, which had an earlier first presentation than uniform items, were recalled significantly more on the first test trial, $F(1,21) = 18.24$, $MSe = 1.12$. This difference favoring expanding items was marginally significant on the second test trial, $F(1,21) = 3.06$, $MSe = 1.46$, and was significant on the third test trial, $F(1,21) = 7.09$, $MSe = 1.16$. The advantage for expanding items, however, disappeared on the final recall test, because a significant number of expanding items were forgotten between the third repetition and the final test, $F(1,21) = 17.01$, $MSe = .34$, and no reliable decrease in performance was found for uniform items.

A significant interaction between spacing and test trials was also found for test-study repetitions, $F(6,126) = 37.89$, $MSe = 1.32$. Expanding items were recalled more than uniform items on the first test trial; the difference was marginally significant, $F(1,21) = 3.12$, $MSe = 1.64$, $p = .09$. This difference between expanding and uniform items, however, disappeared by the second test trial, and was marginally significant in the reverse direction (favoring uniform items) by the third trial, $F(1,21) = 3.84$, $MSe = 1.71$, $p =
Table 3.--Experiment 1a Recall Proportions as a Function of Type and Spacing of Repetitions Across Test Trials

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Test Trial</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Test 1</td>
<td>Test 2</td>
<td>Test 3</td>
<td>Final</td>
</tr>
<tr>
<td>Test-Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanding</td>
<td>.49</td>
<td>.45</td>
<td>.47</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>.32</td>
<td>.37</td>
<td>.36</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>Massed</td>
<td>.88</td>
<td>.85</td>
<td>.84</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Test-Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanding</td>
<td>.42</td>
<td>.59</td>
<td>.68</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>.34</td>
<td>.58</td>
<td>.78</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>Massed</td>
<td>.76</td>
<td>.95</td>
<td>.97</td>
<td>.19</td>
<td></td>
</tr>
</tbody>
</table>
.06. No difference between expanding and uniform items was found for test-study subjects on the final test.

For both test-only and test-study repetitions, massed items were found to be recalled better than distributed (expanding and uniform) items on the first [test-only, \( F(1,21) = 56.72, \text{MSe} = 1.85 \); test-study, \( F(1,21) = 51.78, \text{MSe} = 1.58 \)], the second [test-only, \( F(1,21) = 56.04, \text{MSe} = 2.04 \); test-study, \( F(1,21) = 46.65, \text{MSe} = 2.00 \)], and the third test trials [test-only, \( F(1,21) = 39.17, \text{MSe} = 2.38 \); test-study, \( F(1,21) = 17.39, \text{MSe} = 1.51 \)]. On the final test, however, recall for massed items plummeted, and massed items were recalled significantly less than distributed items [test-only, \( F(1,21) = 17.46, \text{MSe} = .95 \); test-study, \( F(1,21) = 43.73, \text{MSe} = 1.35 \)].

As can be seen in Table 3, the amount of information reacquired across test trials differed for test-only and test-study repetitions. For test-only repetitions, no reliable increase in performance resulted after the first test trial. For test-study items, on the other hand, performance reliably increased from the first to the second trial [expanding, \( F(1,21) = 10.74, \text{MSe} = 1.78 \); uniform, \( F(1,21) = 34.64, \text{MSe} = 1.21 \); massed, \( F(1,21) = 15.4, \text{MSe} = 1.61 \)], and from the second to the third trial [expanding, \( F(1,21) = 7.29, \text{MSe} = .90 \); uniform, \( F(1,21) = 21.93, \text{MSe} = 1.27 \)]. There was a significant amount of forgetting from the third review to the final test for certain spacings of
test-only items [expanding, $F(1, 21) = 17.01$, $MSe = .34$; massed, $F(1, 21) = 143.68$, $MSe = 2.09$] and all spacings of test-study items [expanding, $F(1, 21) = 25.20$, $MSe = 1.17$; uniform, $F(1, 21) = 40.85$, $MSe = 1.45$; massed, $F(1, 21) = 291.49$, $MSe = 1.49$].

Overall, the results of Experiment 1a were very consistent with those of Experiment 1. Repetitions were again found to be highly beneficial, and distributed test-study repetitions were the most beneficial. One difference from the results of Experiment 1 was that expanding test-only trials were not found to be significantly less effective than expanding test-study trials. Although this result provides some support for the claim that expanding test-only trials are an effective method of restudy, expanding test-only trials were not found to be reliably more effective than uniform test-only trials, which is the defining comparison for an expanding test series effect.

Surprisingly, the duration of the initial study period was not found to affect learning. This result is most surprising for test-only and nonrepeated items. In these conditions, the initial presentations are the only study presentations for those items. Perhaps, the 4-s difference between short and long presentations was not large enough to have an effect.

Although the goal of finding an expanding test series effect was again not fulfilled, the tracking of recall
performance across test trials did provide new revealing information concerning the expanding test series effect. It was shown that expanding test-only trials did successfully produce higher levels of recall across repetition tests, but they did not increase final recall performance in comparison to uniform tests. This was the result of greater forgetting between the third repetition test and the final test for items receiving expanding tests. Expanding items may have been forgotten more than uniform items because expanding tests require information to be recalled from long-term memory fewer times than do uniform tests; uniform tests require information to be recalled from long-term memory three times and expanding tests only require information to be recalled from long-term memory twice, assuming that the first test occurs prior to information having been displaced from short-term memory.

Also, it was shown that distributed test-study repetitions provided the learner with the opportunity to acquire more total information across repetitions. Test-only repetitions, on the other hand, only maintained previously learned information in memory. Learners were again found to benefit equally from expanding and uniform test-study trials.

In summary, the combined results of Experiments 1 and 1a provided replications of the test-study retrieval effect and demonstrated large benefits of test-study repetitions
for recall. The results of these experiments, however, failed to provide strong support for either the external validity or the applicability of the expanding test series effect.
CHAPTER IV
EXPERIMENT 2

The purpose of Experiment 2 was to extend the investigation of how and when to restudy information by examining the review of information rather than the repetition of information. As opposed to manipulating the restudy of information during initial learning, as was the case in Experiments 1 and 1a, restudy was manipulated in the form of reviews that were provided immediately after all information had been studied within an initial study session. Initial study in Experiment 2 involved the self-paced presentation of items within a limited period of time. Thus, initial study was different from the single presentation used in the first experiments, since it included multiple repetitions of information. Subsequent reviews were provided at different spacings following initial study as determined by the amount of intervening activity that subjects were asked to complete. Initial study and all subsequent reviews were administered within the same one-hour session.

Method

Subjects. Sixty-six subjects participated from the same subject pool that was used for Experiments 1 and 1a. No participant had been a subject in either of the previous
experiments.

Materials and procedure. The list of to-be-learned items was the same as the list used in Experiments 1 and 1a. The experiment was administered using an IBM-compatible computer. Again, two variables were manipulated: the type of reprocessing (study-only, test-only, and test-study) and the spacing of reprocessing (expanding, uniform, and massed). In contrast to the earlier experiments, the type of reprocessing was manipulated within subjects and the spacing of reprocessing was manipulated between subjects.

Upon arrival, subjects were assigned randomly to one of three spacing conditions. Subjects were first asked to complete the initial learning phase, which was identical for all three conditions. During this initial learning period, all 36 words were available for study within a 400-s period (6 min, 40 s). An item was presented on the screen with the cue above the response and a prompt in the lower left hand corner reading "Press SPACEBAR to see the next item or 'd' to drop that item." That pair remained on the screen until the space bar or letter "d" was pressed. If the space bar was pressed, another item was immediately presented on the screen along with the prompt, but the item remained in the list of to-be-learned items. If "d" was pressed, then that item was removed from the list of items to-be-learned and the next item was presented. The items were presented in cycles such that all items were presented before any item
was re-presented. For each cycle, the items were presented in a random order and that random order was different for each subject. The initial study session was designed to be like studying from a set of flash cards since the subjects were able to circulate through the items as quickly or slowly as they desired and to discard items that had been learned in order to concentrate on the remaining items.

Once the initial study phase was over, subjects were administered a series of three reviews that were separated from one another by having subjects perform a distractor task, which is a vocabulary rating task. The three spacing conditions differed in terms of when the reviews were provided relative to the rating of 14, 14-item distractor lists. Each distractor list took roughly 2 to 3 min to complete. (The administration of distractor lists and reviews is described below.) In the expanding condition, the first review session was provided immediately after initial study; the second review session was provided after three distractor lists had been completed; and the third review session was provided after another six distractor lists had been completed. In the uniform condition, the initial study and review sessions were separated by three gaps of three distractor lists, and in the massed condition, the initial study session and review sessions were presented in succession prior to the administration of all 14 distractor lists. For the expanding and uniform conditions,
the remaining five distractor lists were inserted as a filler between the last review session and the final retention test.

The 14 distractor lists were formed from a random sample of items that were taken from the Oxford American Dictionary (see D'Anna, Zechmeister, & Hall, 1991). Each distractor list consisted of 14 total items. For each 14-word list, the words were randomly presented one at a time. Each word first appeared alone for 1.5 s, followed by the prompt: "Please rate your knowledge of this word." A scale also appeared: "1-I have never seen this word before, 2-I have seen this word but do not know its definition, 3-I have seen this word and have some idea of its meaning, 4-I know this word well enough to recognize its meaning, or 5-I know this word well enough to define it." The word and the scale appeared on the screen for 15 s or until a rating was made, whereupon the next item was presented. There was a distinct beginning and end to each distractor list.

For each review session, the critical items were divided randomly into four sets of eight items and assigned to one of four conditions: study-only, test-only, test/study, or no review. Items were assigned randomly to each of the conditions. In the study-only condition, the cue as well as the response were presented to the subject for a 12-s period. In the test-only condition, the cue was presented without the response for the 12-s period; a cursor box also
was presented, and subjects were prompted to attempt to type the appropriate response and press return. If an answer was registered before the 12-s period was over, a message appeared letting the subject know that the answer had been registered, but the next item was not presented until the entire 12-s period had expired. In the test/study condition, the cue was presented alone with a cursor box for the first 8 s wherein subjects were asked to enter an appropriate response; this was followed by a 4-s period where the cue and response were presented together. Again, if an answer was registered prior to the completion of the initial 8-s test period, a message verified the registration, but the study period did not begin until the entire 8-s test period was over. Control items (no review) were not presented in any of the review sessions.

Items remained in their respective conditions (study-only, etc...) for all review sessions. Within a review session, the order of item presentation was completely random for each subject, and the order was randomly shuffled before each review session. Once all review sessions and distractor lists had been administered, the final recall test was administered just as in Experiment 1.

**Results and Discussion**

The results for the final cued-recall test of Experiment 2 were analyzed using a 4 X 3 mixed ANOVA design. The proportions of correct responses are summarized in Table 4.
As expected, the type of review significantly affected recall performance, $F(3,189) = 105.41$. Test-study items were recalled more than study-only items, $F(1,65) = 31.09$, $MSe = 1.06$; study-only items were recalled more than test-only items, $F(1,65) = 64.66$, $MSe = 2.43$; and test-only items were recalled more than nonreviewed items, $F(1,65) = 5.05$, $MSe = 1.64$. Spacing of review, however, did not significantly affect recall performance and did not interact with type of review.

An additional analysis was conducted to track the difference in performance between test-study and test-only items across the three review test trials and the final test trial. A 3 (spacing) X 2 (type) X 4 (trials) mixed design ANOVA was used to analyze these results, which are summarized in Table 5. Because spacing did not have a significant overall effect and did not interact with either of the other variables, the results, as illustrated in Figure 1, were combined for all spacing conditions. As can be seen in Figure 1, main effects were found for both type of review, $F(1,63) = 50.62$, $MSe = 5.99$, and test trials, $F(3,189) = 160.28$, $MSe = .60$, and these variables were found to interact with each other, $F(3,189) = 136.04$, $MSe = .57$. For test-only items, performance remained generally consistent across trials; only the increase in performance from the third review to the final test was reliable, $F(1, 65) = 12.60$, $MSe = .14$. For test-study items, on the other
Table 4.--Experiment 2 Recall Proportions as a Function of Type and Spacing of Immediate Reviews

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Type</th>
<th>Nonreviewed</th>
<th>Test-Only</th>
<th>Study-Only</th>
<th>Test-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanding</td>
<td>.26</td>
<td>.34</td>
<td>.56</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>.34</td>
<td>.35</td>
<td>.65</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Massed</td>
<td>.22</td>
<td>.32</td>
<td>.62</td>
<td>.72</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.---Experiment 2 Recall Proportions as a Function of Type and Spacing of Immediate Reviews Across Test Trials

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Test Trial</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Review 1</td>
<td>Review 2</td>
<td>Review 3</td>
<td>Final</td>
</tr>
<tr>
<td>Expanding</td>
<td>.30</td>
<td>.30</td>
<td>.31</td>
<td>.34</td>
</tr>
<tr>
<td>Uniform</td>
<td>.34</td>
<td>.34</td>
<td>.32</td>
<td>.35</td>
</tr>
<tr>
<td>Massed</td>
<td>.28</td>
<td>.29</td>
<td>.30</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>Test-Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanding</td>
<td>.28</td>
<td>.41</td>
<td>.53</td>
<td>.70</td>
</tr>
<tr>
<td>Uniform</td>
<td>.26</td>
<td>.45</td>
<td>.61</td>
<td>.78</td>
</tr>
<tr>
<td>Massed</td>
<td>.25</td>
<td>.46</td>
<td>.61</td>
<td>.72</td>
</tr>
</tbody>
</table>
Fig. 1. Experiment 2 Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials
hand, performance was found to increase steadily from the first to the second review trial, $F(1,65) = 83.56$, $MSe = .77$, from the second to the third review trial, $F(1,65) = 52.64$, $MSe = .88$, and from the third review trial to the final test, $F(1,65) = 92.51$, $MSe = .51$.

The goal of Experiment 2 was to extend the findings of Experiments 1 and 1a, by examining how and when to review information. As was the case for repetitions, reviews were found to increase recall performance, and test-study reprocessing, again, was most effective. The superiority of test-study review, at least over test-only review, was traced to a steady increase in learning across review trials that was absent for test-only items. These results represent the first demonstration of a retrieval effect that used a distributed study-only review control group. This demonstration was limited, however, to test-study review; test-only review was found to be reliably less effective than study-only review. Thus, the need for additional study during reprocessing appears to have been more important for reviews than it was for repetitions because of the longer interval between initial study and restudy.

One unexpected finding of Experiment 2 was the lack of any effect for the spacing of reviews. It appears that the spacings, which differed from each other in terms of minutes of distractor activity, were not sufficiently different from each other. Thus, neither a spacing effect nor an expanding
test series effect was found for reviews in Experiment 2. Whether these results would hold when reviews were spaced across days, was the question addressed in Experiments 3 and 3a.
CHAPTER V
EXPERIMENT 3

The investigation of how type and spacing of review influence learning was continued in Experiment 3. In contrast to Experiment 2, however, reviews in Experiment 3 were spaced across days. This was done to increase the differences between spacing schedules. This design also had the benefit of examining "...retrieval schedules over time periods that are relevant to the interests of educators" (p. 321), which is a goal that has been suggested to memory researchers (Bahrick et al., 1993).

Method

Subjects. A total of 54 students from the subject pool used in the previous experiments and from two nonintroductory psychology courses at Loyola University Chicago participated. Only 43 of those who started the experiment successfully completed the experiment; four subjects, three subjects, and four subjects, respectively, were lost from the expanding, uniform, and massed spacing conditions. No subject had participated in any of the earlier experiments.

Materials and procedure. Pilot testing suggested that when learning was examined across a 9-day period, a number of changes in the stimulus items and initial learning proce-
dures needed to be made in order to prevent extremely low recall levels. First, to involve subjects in the task to a greater extent, a list of uncommon but real vocabulary items was used (McDaniel \& Pressley, 1984). An example item is *handsel-payment*. Second, unlike the procedure followed in the previous experiments, only the initial learning phase was computerized in Experiment 3.

As in Experiment 2, spacing of review was manipulated as a between-subjects variable and type of review was manipulated as a within-subjects variable. Upon arrival, subjects were randomly assigned to one of three spacing conditions. Subjects in all conditions first completed an initial learning phase. In this initial learning period, subjects were first presented 40 vocabulary words using the computerized flash-card study procedure used in Experiment 2. Unlike Experiment 2, there was no time limit for the initial study period. Study continued until all of the items had been dropped out of the to-be-studied list. Then all of the cue members of the items were presented along with a cursor box and a test prompt asking subjects to type the appropriate definition or response member for each word. If subjects answered correctly, the program indicated the response was correct and the item was dropped from the list; if subjects answered incorrectly, the correct response was provided and that item was placed in a set of items that would be re-presented. Once all items had been tested,
those items that were answered incorrectly continued to be
restudied and retested in new random orders until the sub-
ject answered each item correctly one time.

Subjects within each of the spacing conditions then
received review sessions and rating sessions at different
spacings. Each review session was administered using a
flash-card booklet that was prearranged to provide, in a
random order, study-only review (one card with a word and
definition) for 10 items, test-only review (one card with
just a word) for 10 items, test-study review (one card with
just a word followed by a second card with the same word and
definition) for 10 items, and no review for a remaining 10
items. Four different booklet orders were created so that
across subjects each item served in each of the within-
subjects conditions equally often. Subjects reviewed the
items at their own pace, but they were asked not to go
backwards through the booklet at any time. For test items,
subjects were asked to write the appropriate definition on
the test card. Rating sessions presented the same pool of
distractor items as presented for Experiment 2, using three
different rating sheets comprised of 66, 66, and 67 items
respectively. Rating sessions were used in this experiment
to balance the number of days that participation was re-
quired for each of the spacing conditions to prevent any
selective loss of subjects; if rating sheets were not used,
subjects in the massed condition would only have participat-
ed on 2 days and subjects in the expanding and uniform conditions would have participated on 5 days.

Subjects in the expanding spacing condition were asked to complete all three rating sessions immediately after initial study. Then, following a 1-day delay, subjects completed the first review booklet, followed by the second review booklet 2 days after that, and the final review booklet 3 days after that. Subjects were asked to complete the review booklets at any time within the designated day. Subjects' completion of the review booklets was not monitored. A final retention test that was monitored was given 3 days after the final review session. The same procedure was used for the uniform spacing condition except that all review sessions were spaced at 2-day intervals with the final retention test given 3 days after the final review. The massed spacing condition differed from the other two spacing conditions in that all the three review sessions were administered immediately after initial study. So that massed subjects would also have to complete parts of the experiment on their own, they were asked to complete the three rating sessions according to either an expanding or uniform spacing. Expanding ratings were spaced at expanding gaps of 1, 2, and 3 days, and uniform gaps were spaced at gaps of 2 days each. Roughly, an equal number of subjects within the massed condition received expanding and uniform spacings of ratings.
In all spacing conditions, subjects were given the final cued-recall test 9 days after the initial study session. The final test was administered in the laboratory using a pencil and paper procedure. Cue-members for all critical items were arranged in the same random order for all subjects. Subjects were given as much time as they needed to write in a definition for each of the words.

Results and Discussion

As was mentioned previously, 11 subjects did not successfully complete the experiment. This was because they either failed to complete a review or rating session on the scheduled day or they failed to return for the final test. An additional subject within the uniform condition was randomly excluded from all analyses to provide an equal number of 14 subjects within each of the conditions.

The results for the final cued-recall test of Experiment 3 are summarized in Table 6. A 4 X 3 mixed design ANOVA was used to analyze these results. A significant interaction was found between the spacing and type of review, $F(6, 117) = 7.85$, $MSe = 2.88$, in addition to significant main effects for each variable: spacing, $F(2, 39) = 13.41$, $MSe = 11.78$, and type, $F(3, 117) = 91.18$, $MSe = 2.88$.

A closer look at the interaction between type and spacing of review, reveals that type of review showed a similar effect when reviews were spaced at expanding and uniform intervals, $F(3, 78) < 1$, $MSe = 2.68$. For expanding and
Table 6—Experiment 3 Recall Proportions as a Function of Type and Spacing of Delayed Reviews

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Type</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonreviewed</td>
<td>Test-Only</td>
<td>Study-Only</td>
<td>Test-Study</td>
</tr>
<tr>
<td>Expanding</td>
<td>.18</td>
<td>.55</td>
<td>.78</td>
<td>.84</td>
</tr>
<tr>
<td>Uniform</td>
<td>.27</td>
<td>.64</td>
<td>.91</td>
<td>.98</td>
</tr>
<tr>
<td>Massed</td>
<td>.14</td>
<td>.50</td>
<td>.34</td>
<td>.49</td>
</tr>
</tbody>
</table>
uniform spacings, test-only items were recalled more than nonreviewed items, $F(1,27) = 49.16$, $MSe = 3.85$, study-only items were recalled more than test-only items, $F(1,27) = 36.07$, $MSe = 2.43$, and test-study items were recalled more than study-only items, $F(1,27) = 4.44$, $MSe = 1.30$. It should also be noted that subjects receiving a uniform spacing of review recalled slightly more overall than subjects receiving expanding review, $F(1,26) = 3.85$, $MSe = 9.22$, $p = .061$.

For massed review, a different pattern of results emerged. No difference was apparent between the recall of test-study and test-only items, and the recall of study-only items was significantly less, $F(1,13) = 6.47$, $MSe = 2.07$. It is noteworthy that massed review was very similar to the immediate review conditions used in Experiment 2 except for the longer retention interval; yet, study-only review was more effective than test-only review in Experiment 2 and test-only review was more effective than study-only review in Experiment 3. Study-only items were, however, still recalled more than nonreviewed items, $F(1,13) = 5.51$, $MSe = 4.73$.

Looking at the interaction from a different perspective, spacing was only found to affect performance for test-study, $F(2,39) = 15.57$, $MSe = 5.60$, and study-only items, $F(2,39) = 21.29$, $MSe = 5.91$. For test-study items, a spacing effect was found between massed and expanding items,
\( F(1,39) = 14.69, \text{MSe} = 5.60 \), but an expanding testing effect was not apparent, as expanding and uniform items were recalled equally well. A spacing effect was also found for study-only items, \( F(1,39) = 23.23, \text{MSe} = 5.91 \), and there was no reliable difference between expanding and uniform spacings of study-only trials.

As in the previous experiments, additional analyses were conducted using a 2 (type) X 3 (spacing) X 4 (trials) mixed ANOVA to trace the difference between test-study and test-only items (see Table 7). Because some subjects did not write responses in their booklets, only 33 subjects were included in this analysis. The three-way interaction between type, spacing, and review trials was found to be marginally significant, \( F(6,90) = 2.01, \text{MSe} = .915, p = .073 \); thus, the effects of type and trials were analyzed for expanding, uniform, and massed spacings separately. Again, expanding and uniform spacings were found to produce the same basic result pattern, \( F(3,60) = 1.21, \text{MSe} = .58, p > .10 \), which was different from the pattern of results found for massed items. One major difference was the presence of a type by trials interaction for expanding and uniform spacings combined, \( F(3,60) = 3.73, \text{MSe} = .58 \), and the absence of this interaction for massed spacings. The type by trials interaction for expanding and uniform items combined is illustrated in Figure 2. For expanding and uniform spacings, recall was shown to increase across reviews for test-
Table 7.—Experiment 3 Recall Proportions as a Function of Type and Spacing of Delayed Reviews Across Test Trials

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Test Trial</th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test-Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanding</td>
<td>.65</td>
<td>.65</td>
<td>.65</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>.65</td>
<td>.68</td>
<td>.64</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>Massed</td>
<td>.85</td>
<td>.84</td>
<td>.84</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test-Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanding</td>
<td>.85</td>
<td>.87</td>
<td>.94</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>.80</td>
<td>.95</td>
<td>.94</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>Massed</td>
<td>.91</td>
<td>.89</td>
<td>.89</td>
<td>.49</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2. Experiment 3 Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials
study items, $F(3,63) = 4.59$, $MSe = 1.26$, and to remain unchanged across reviews for test-only items. As can be seen in Figure 2, the increase for test-study items is largely the result of the jump in performance from review 1 to review 2, $F(1,21) = 11.26$, $MSe = .73$.

Unexpectedly, recall was significantly higher for test-study items than for test-only items on the first review test, $F(1,21) = 11.26$, $MSe = 2.91$, where there should be no difference between these conditions since the manipulation had yet to be employed. The probability of this difference happening by chance alone is quite low, $p = .003$. Thus, it would appear that some subjects must have cheated on this review test by looking forward in their review booklet when responding to test-study tests. Although this finding clearly compromises the integrity of these results, it is important to note that only three data points were possibly affected: reviews 1 to 3 for test-study items. There was no way for subjects to cheat on the final recall test. Thus, subjects did legitimately rise to the high level of retention found for test-study items by the final cued-recall test.

For the massed review spacing, no differences were apparent for test-study and test-only reviews across review trials. It was shown, however, that for both test-study and test-only tests performance dramatically decreased from the third review to the final cued-recall test, $F(1,10) = 27.06$, 
Overall, Experiment 3 was successful in demonstrating both the overall effectiveness of review and the superiority of test-study review for reviews spaced across a 9-day period. Thus, to this point, benefits of restudy have been shown for repetitions, immediate reviews, and delayed reviews. Moreover, test-study restudy has consistently been the most effective type of restudy.

Also in Experiment 3, a reversal was found between the recall levels of massed study-only and massed test-only items from the recall levels found in Experiment 2, despite the fact that review was given immediately after initial study in both situations. In Experiment 3, test-only items were better recalled after the 9-day retention interval that was used. This finding suggests that study-only review is more vulnerable to forgetting than test-only review.

Unlike Experiment 2, the spacing of review was found to affect performance in Experiment 3. A spacing effect was apparent for both test-study and study-only items; expanding and uniform items were recalled more than massed items. Because the interval from initial study to the final test was kept constant across spacings, massed items did have a longer final retention interval than expanding or uniform items. Thus, it is possible that considerable forgetting would similarly be found for expanding and uniform items if they endured a retention interval of more than a week. This
possibility was examined in Experiment 3a.

An expanding test series effect was not found for test-only, test-study, or study-only items in Experiment 3. The expanding test series effect, thus, remains undemonstrated for restudy in the form of reviews. The applicability of expanding tests for classroom learning is clearly limited if the expanding test series effect cannot be extended to reviews or test-study trials.
CHAPTER VI

EXPERIMENT 3A

The results of Experiment 3 were encouraging in that information that received distributed review was remembered substantially better than information that was not reviewed or was reviewed immediately after initial study. However, it is possible that the positive effects of review were limited to the 3-day retention interval that was used in Experiment 3. Therefore, a longer retention interval of 8 days was employed in Experiment 3a, in order to test the durability of the review effects found in Experiment 3.

It was expected that the recall of study-only items would decrease in relation to test-only and test-study items in Experiment 3a. This pattern of results was suggested in the data obtained from the subjects who were excluded from Experiment 3, because they did not complete their final retention test until after the designed 3-day delay. It also made sense that the greater effort demanded by tests would maintain information in memory longer than additional study would.

Method

Subjects. A total of 28 students who were enrolled in an upper-division psychology course at Loyola University
Chicago participated in the experiment as part of a course requirement. No subject had participated in any of the earlier experiments.

**Materials and procedure.** The materials and procedures for Experiment 3a were very similar to those of Experiment 3. There were some differences, however. First, due to the limited number of subjects, only two spacings of review were examined, expanding and uniform. Second, although the spacings between review sessions were identical to those used in Experiment 3 (1 day, 2 days, and 3 days for expanding, and 2 days, 2 days, and 2 days for uniform), the delay between the final review session and the final test was increased to 8 days. Finally, in an attempt to prevent subjects from cheating during review tests, the procedure for review sessions was modified. Rather than providing booklets that randomly presented the 10 study-only items, the 10 test-only items, and the 10 test-study items, these items were presented using two full sheets of paper: a test-sheet and a study-sheet. On the test-sheet, the cue members of the test-only and test-study items were presented in a mixed order with a blank line presented after each cue. Subjects were instructed to provide a definition for as many of the words as they could. Stapled closed to the back of the test-sheet was the study-sheet. Subjects were explicitly told not to write on the test-sheet once the study sheet had been opened. On the study-sheet, the cue words with
their definitions were presented in a mixed order for the study-only and test-study conditions. The same words were presented as study-only, test-only, and test-study items on each of the three reviews, but the order of presentation changed. As in Experiment 3, subjects were presented all 40 words that had been originally learned (30 reviewed, 10 not reviewed) on the final paper-and-pencil, cued-recall test. The same random order of items was used for all subjects, and subjects were given as long as they needed to complete the test.

Results and Discussion

Results for the final cued-recall test of Experiment 3a are summarized in Table 8. A 4 X 2 mixed ANOVA was used to analyze these results. No difference in overall level of recall was apparent between the expanding and uniform spacing conditions. Moreover, spacing did not interact with type of review to affect performance. A significant effect was found for type of review, $F(3,78) = 55.24$, $MSe = 3.77$. As was the case for Experiment 3, test-study review resulted in higher performance than study-only review, $F(1,27) = 19.59$, $MSe = 3.97$, and no review resulted in the lesser recall than test-only review, $F(1,27) = 45.44$, $MSe = 5.02$. As was expected, the level of recall for study-only review dropped relative to the level of recall for test-study and test-only review. In fact, there was no reliable difference in performance between study-only review and test-only
Table 8.--Experiment 3a Recall Proportions as a Function of Type and Spacing of Delayed Reviews

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Type</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonreviewed</td>
<td>Test-Only</td>
<td>Study-Only</td>
<td>Test-Study</td>
</tr>
<tr>
<td>Expanding</td>
<td>.19</td>
<td>.51</td>
<td>.60</td>
<td>.82</td>
</tr>
<tr>
<td>Uniform</td>
<td>.20</td>
<td>.69</td>
<td>.64</td>
<td>.89</td>
</tr>
</tbody>
</table>
Additional analyses were conducted, using a 2 (type) X 2 (spacing) X 4 (trials) mixed ANOVA, in order to trace the difference in performance between test-study review and test-only review across testing opportunities. A summary is found in Table 9. Expanding review and uniform review yielded the same pattern of results, $F(3,78) < 1$, $MSe = 1.24$. Thus, the results were combined for expanding and uniform spacings in Figure 3. As can be seen in Figure 3, test-study review and test-only review showed different patterns of results across test trials, $F(3,78) = 10.13$, $MSe = 1.24$. For test-only review, performance remained largely unchanged across the three review tests and final test, $F(3,81) = 1.09$, $MSe = .71$, $p > .10$. For test-study review, performance significantly increased across test trials, $F(3,81) = 14.42$, $MSe = 1.47$. This increase was the result of the jump in performance from the first to the second review test, $F(1,27) = 22.20$, $MSe = 2.18$. Although the highest performance level was reached on the second review test, no reliable performance decrement was found for the third review and final tests.

In contrast to Experiment 3, there was no reliable difference in performance between test-study and test-only items on the first review test, although the means did favor test-study review. Significant differences favoring test-study review were found for all remaining tests: review 2,
Table 9.--Experiment 3a Recall Proportions as a Function of Type and Spacing of Delayed Reviews Across Test Trials

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Test Trial</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Review 1</td>
<td>Review 2</td>
<td>Review 3</td>
</tr>
<tr>
<td>Test-Only</td>
<td></td>
<td>.55</td>
<td>.56</td>
<td>.56</td>
</tr>
<tr>
<td>Expanding</td>
<td></td>
<td>.73</td>
<td>.70</td>
<td>.69</td>
</tr>
<tr>
<td>Uniform</td>
<td></td>
<td>.73</td>
<td>.94</td>
<td>.94</td>
</tr>
<tr>
<td>Test-Study</td>
<td></td>
<td>.71</td>
<td>.87</td>
<td>.86</td>
</tr>
<tr>
<td>Expanding</td>
<td></td>
<td>.74</td>
<td>.94</td>
<td>.94</td>
</tr>
<tr>
<td>Uniform</td>
<td></td>
<td>.74</td>
<td>.94</td>
<td>.94</td>
</tr>
</tbody>
</table>
Fig 3. Experiment 3a Recall Proportions for Distributed Items as a Function of Type of Review and Test Trials
The results of Experiment 3a provided a needed extension of the results from Experiment 3. It was again shown that review was an effective tool for the maintenance of knowledge in memory. This was especially true for review with testing, since no appreciable forgetting was observed for test-only and test-study items when the delay between the third review and final recall was increased to 8 days in this experiment.

Study-only items, however, did not maintain their effectiveness relative to test-study and test-only items when the retention interval was increased to eight days. Study-only items were now recalled to the same degree as test-only items, a level far beneath that of test-study recall. This difference in performance for study-only items across Experiments 3 and 3a, suggests that study-only review is more vulnerable to forgetting than test-study or test-only review. Thus, test-study review has proven to be most effective in comparison to the other review methods, within the learning situation of greatest applicability for educators.

An expanding test series effect was again not found in Experiment 3a. Combined, an expanding test series effect was not demonstrated in three attempts using reviews and in two attempts using repetitions. These results suggest that
repetition/review effects, spacing effects, and test-study retrieval effects are all of greater reliability than the expanding test series effect. The implications of these results and suggestions for future research are discussed in the next chapter.
Before addressing how and when information was restudied best, the question of whether restudy benefitted learning will be addressed. Indeed, the results of the present experiments replicated the results of previous research by demonstrating that repeated items were recalled better than nonrepeated items (Experiments 1 and 1a), that immediately reviewed items were recalled more than nonreviewed items (Experiment 2), and that delayed reviewed items were recalled more than nonreviewed items (Experiments 3 and 3a). Although these results were definitely expected, the sizes of these differences were somewhat surprising. In Table 10, the proportions of correct responses were combined for all types of restudy and contrasted with control items that were not restudied. As can be seen in Table 10, the level of recall for repeated items was roughly three times that for nonrepeated items (Experiments 1 and 1a), the level of recall for immediately reviewed items was more than twice that for nonreviewed items (Experiment 2), and the level of recall for delayed review items was more than three times that for nonreviewed items (Experiments 3 and 3a).

Thus, the present results add to the fairly large
Table 10.---A Comparison of Reprocessed and Nonreprocessed Items Across Experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Reprocessed</th>
<th>Nonreprocessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>.42</td>
<td>.15</td>
</tr>
<tr>
<td>Experiment 1a</td>
<td>.31</td>
<td>.09</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>.56</td>
<td>.27</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>.67</td>
<td>.20</td>
</tr>
<tr>
<td>Experiment 3a</td>
<td>.69</td>
<td>.20</td>
</tr>
</tbody>
</table>
number of studies reporting large benefits of restudy. Although many people are aware of the benefits of restudy, many fail to utilize this benefit. For example, often educators only have time to present material one time throughout a semester or students only study a given topic one time in preparation for a test. The present results would suggest that in these situations the potential for learning is minimized.

How Should Information be Restudied?

Assuming that information is restudied, it is of central importance to know how to restudy. The question of how to best restudy information was investigated by comparing three different types of restudy: test-study, test-only, and study-only restudy. Although all three types of restudy were found to be more effective than not restudying, these methods were not found to be equally effective. The question of how information was restudied best will be addressed separately for repetitions and reviews. This will be followed by a discussion of when information should be restudied, which also will be addressed for repetitions and reviews.

Repetitions. In Experiments 1 and 1a, test-study repetitions were generally found to be the most effective type of restudy. Specifically, the highest levels of recall were found for test-study repetitions that had been distributed. These results replicate those of Carrier and Pashler.
(1992), demonstrating that retrieval during restudy promotes learning in ways that study-only repetitions cannot.

As was mentioned in the introduction, there are a number of possible reasons why test-study repetitions were more effective than study-only repetitions. First, test-study trials may have simply required greater effort and thus provided more effective processing at the time of acquisition. Second, the process of successfully retrieving an item from long-term memory may have strengthened the successful retrieval route, making that item less vulnerable to forgetting. Finally, test-study trials may have increased the learners' awareness of whether an item had been learned or whether an elaborator was working, and learners may have increased effort to learn or have generated more successful elaborators for nonretrievable items.

The analysis in Experiment 1a of test-study trials across repetition tests provided some insight concerning the need for feedback in order for retrieval to be effective. In the test-only condition, where subjects were not provided with feedback as to whether their responses were correct, recall levels generally remained constant across repetitions and decreased slightly on the final recall test. Final recall performance for test-only repetitions was not reliably higher than that for study-only repetitions. However, in the test-study condition, where feedback was given, recall levels increased across repetition tests, which
resulted in a higher final level of recall than for study-only repetitions. These results demonstrate that feedback is necessary for learners to reacquire forgotten information.

Experiments 1 and 1a also compared test-study repetitions and study-only repetitions for massed items, which is a comparison that was not made by Carrier and Pashler (1992). For massed items, test-study repetitions were not found to be reliably more effective than study-only repetitions. Most likely, this occurred because information never left short-term memory before it was tested. The high level of recall for massed items across repetition tests supports this possibility. Thus, retrieving information from memory presumably did not require greater effort, did not strengthen retrieval routes from long-term memory, and did not alert the learner as to whether initial learning was adequate.

The practical implication of the greater effectiveness of test-study repetition is quite simple. When information is repeated within a study session, learners should be provided with test-study repetitions that are spaced apart from each other so that the tests will be productive. This could be easily done, for example, for a learner interested in studying facts, such as the capitals of the fifty states, by using flash cards. Similarly, during lectures, questions directed to the students about already presented information should be effective, assuming both that enough time has
passed for the test to be challenging and that feedback is given following the test.

**Reviews.** Experiments 2, 3, and 3a, provided the first comparisons of test-study reviews and study-only reviews. Extending the results found for repetitions, test-study review was found to be reliably more effective than study-only review when the review of information occurred immediately after initial study (Experiment 2). Test-study review was also generally shown to be more effective than study-only review for delayed review (Experiments 3 and 3a). These results are important because they demonstrate the beneficial effect of retrieval for reviews of information, which are of great interest to educators.

The reasons why test-study review was effective are probably very similar to the reasons why test-study repetitions were effective. However, there were some differences between review tests and repetition tests that must be taken into account. For instance, because review tests were given after a longer retention interval than repetition tests, retrieval was more difficult. As a result, the need for additional study opportunities or feedback for tests became more important. Thus, test-only reviews elicited lower levels of performance than study-only reviews, since little improvement could be made after the first review test for test-only items. For test-study and study-only items, on the other hand, forgotten information could continue to be
re-acquired during review sessions. Again, recall of test-study items may have been more successful than that of study-only items because test-study review required greater effort during processing, strengthened successful retrieval routes, or alerted subjects to use more effective elaborators.

Whatever the mechanisms were that made test-study trials more effective, the results of Experiments 3 and 3a suggested that the benefits of test-study review were more durable than those of study-only review. For example, in Experiment 3, when the retention interval was three days, distributed study-only items (.85) were recalled slightly less than distributed test-study items (.91) and significantly more than distributed test-only items (.60). However, in Experiment 3a, when the retention interval was increased to eight days, test-study items (.85) and test-only items (.60) showed little decrease in recall, but study-only items (.62) showed a large decrease. Study-only items now trailed test-study items by twenty-three percentage points as opposed to six, and no significant difference was found between study-only and test-only items. Similarly, in Experiment 2 study-only review (.61) was more effective than test-only review (.34), but in the massed condition of Experiment 3 test-only review (.50) was more effective than study-only review (.34). The massed condition of Experiment 3 was similar to the spacing conditions used in Experiment
2, except that the final test occurred after a longer, nine-day retention interval. Thus, it would appear that tests serve to effectively maintain information in memory; whereas, additional study allows for the reacquisition of lost information, but this information is not maintained as long in memory.

The greater effectiveness of test-study review has obvious implications for education. Test-study reviews provided after initial study or across days are effective learning techniques. For the long-term retention of information, distributing review by using test-study review appears to be most effective, as the high level of retention found for distributed items was shown to hold across retention intervals longer than a week. Thus, providing quizzes on information (with feedback) in class on a regular basis should greatly improve students' learning of the tested information.

When Should Information be Restudied?

In addition to knowing how information should be restudied, it is also important to know when to restudy. The question of when was primarily addressed in the present study by manipulating the spacing of restudy. This question will again be answered separately for repetitions and reviews.

Repetitions. The spacing schedules within Experiments 1 and 1a were shown to significantly affect performance.
When subjects were given distributed spacing schedules (expanding or uniform spacings) in Experiments 1 and 1a, significantly more items were recalled than when subjects were given massed spacings. This spacing effect, as expected, produced large differences in performance between massed (.17) and distributed items (.375). The spacing effect was largest for test-study repetitions, was second largest for test-only repetitions, and was smallest for study-only items. This finding is interesting since the smallest effect was found for study-only restudy and this method of restudy is used most often in demonstrations of the spacing effect.

Also of interest, was whether expanding and uniform distributions of restudy were differentially effective. Past research had shown expanding test-only repetitions to be more effective than uniform test-only repetitions (e.g., Landauer & Bjork, 1978; Shaughnessy et al., in preparation). In Experiment 1, expanding presentations were not shown to be more effective than uniform presentations for test-only trials, test-study trials, or study-only trials. In Experiment 1a, more filler items were added and the length of the initial presentation of items was manipulated in an attempt to find this effect. Unfortunately, length of initial presentation did not reliably affect learning, and again no significant differences were found between expanding and uniform spacing schedules.
It is surprising that in Experiment 1a the length of initial presentation did not affect performance. It had been expected that when the amount of time given for initial study was cut in half that subsequent retrieval would be more difficult, especially for test-only repetitions where the initial presentation was the only study presentation. Performance was not different, however, for long and short study; perhaps, for the discrete verbal items used in this experiment, connections between words either occur quickly or not at all.

In Experiment 1a, subjects' responses on test repetitions proved to be very informative concerning the lack of an expanding test series effect. As can be seen in Table 3, the first expanding test, spaced one item away from initial study, was successful in eliciting a higher recall level than the first uniform test, spaced five items away. This difference was significant for test-only trials and marginally significant for test-study trials. For test-only items, this advantage for expanding tests continued across the second and third gaps, which were of five items and nine items, respectively, for expanding items, and five items both times for uniform items. This advantage, however, did not carry over to the fourth test, the final cued-recall test. Although an expanding schedule allowed more items to be maintained across the first three tests, significant numbers of these items were forgotten on the final
test. Uniform spacings, on the other hand, made it more difficult for items to be recalled on the first three tests, but no significant decrease in performance was found from the third repetition to the final test. Thus, it would appear that uniform and expanding spacings provide slightly different benefits, that, at least in the present experiments, were equally effective.

Explanations of why expanding and uniform items provided different benefits can be found in the number of times an item was retrieved from long-term memory and in the selection of items. In the expanding condition, the first repetition was intended to provide a bridge so that information could be retrieved on subsequent repetitions. Because the first repetition test was placed closely to the initial presentations of items, the information may not have been displaced from short-term memory at the time of the first repetition test, and thus some more difficult items could have been successfully recalled. Although this bridging increased the likelihood that recall would be successful on the subsequent repetition tests, it also eliminated one attempt to recall information from long-term memory and it allowed for some more difficult items to have been successfully recalled on the repetition tests. Uniform tests, on the other hand, provided three opportunities for information to be recalled from long-term memory and created a less difficult pool of successfully recalled items because the
repetition tests were more difficult. Thus, it follows that expanding items would be more vulnerable to forgetting since they were recalled less times from long-term memory and they were more difficult items to remember.

Absolutely no evidence was found in support of the expanding test series effect for test-study repetitions. When expanding and uniform test-study repetitions were compared, a marginally significant difference in performance was found favoring expanding items on the first test trial. This advantage, however, was not maintained across test trials, as uniform and expanding items were recalled similarly or uniform items were recalled more on the three remaining tests.

The results of Experiment 1 and 1a question the claim that expanding testing is "a very powerful strategy that is easy to use and widely applicable,...probably more broadly useful than any of the more traditional visual imagery mnemonics" (Baddeley, 1990, p. 158). The present results suggested that the expanding test series effect is unreliable for test-only trials and does not generalize to test-study trials. If expanding tests are only effective on an inconsistent basis in situations where feedback is not provided after tests, their applicability is clearly very limited. Perhaps this would be a method to try when learning names at a cocktail party where one does not want to ask people their names more than once, but when learning facts
in situations where feedback can be obtained, any distribution of test-study trials apparently would be more effective.

Reviews. The spacing of immediate review was unexpectedly not found to affect recall performance in Experiment 2. However, the low level of recall on the first review test for massed items in Experiment 2 suggests that information, when it was recalled, was being retrieved from long-term memory. Thus, unlike massed repetitions (Experiments 1 and 1a), massed review had distributed practice built in to some extent in Experiment 2. Although expanding and uniform reviews were designed to provide wider distribution and more effective practice, it appears that the differences in spacings within a single 1-hr session were not sufficient to differentially influence performance.

When the reviews were spaced across days instead of minutes, the spacing of review was found to be significant (Experiment 3). As was the case for repetitions, expanding and uniform spacings were found to be more effective than massed reviews. These effects, however, were limited to test-study and study-only review. Although the retention interval was longer for massed reviews than for expanding or uniform reviews in Experiment 3, the level of recall for expanding and uniform reviews did not decrease substantially in Experiment 3a, which used a retention interval comparable to that used for massed reviews in Experiment 3. These
results make it doubtful that the superiority of spaced reviews over massed reviews was related to the longer retention interval for massed items in Experiment 3. Thus, it can be concluded that distributing restudy is very effective for delayed reviews as well as repetitions.

An expanding schedule of review was not found to be more effective than an uniform schedule for immediate or delayed reviews. However, the spacing differences between expanding and uniform reviews may not have been large enough in these experiments to produce expanding test series effects. For example, in Experiments 3 and 3a the first review was given one day following initial learning for expanding review and two days following initial learning for uniform items. As can be seen in Tables 7 and 9, this 1-day difference did not result in greater recall on the first test for the expanding condition. In future research, a greater difference in spacing between the first reviews for expanding and uniform spacings should be used to see if a recall advantage could be developed for expanding items.

From a practical perspective, the manufacturing of an expanding test series effect for review tests will still be of limited applicability for education unless it is found for test-study reviews. As can be seen in Figures 1, 2, and 3, test-study items continued to be learned across review sessions regardless of whether they were recalled on the first review test. Thus, shaping the tests to generate
successful recall seems to have been of little benefit. All that appears important for test-study review is that the reviews are spaced far enough from each other so that they are challenging.

Repetitions Versus Reviews

The question of when information should be restudied can also be addressed by asking whether repetitions, immediate reviews, or delayed reviews are most beneficial. Although repetitions, immediate reviews, and delayed reviews were all found to be beneficial, reviews appear to be most beneficial, at least for long-term retention. First, as can be seen in Figures 1 through 3, the effect of repetitions is not as durable as the effect of reviews. For repetitions, a sizable amount of forgetting was found from the third repetition to the final test; whereas, for reviews, no reliable differences in forgetting were found.

Second, because of changes in the experimental procedure, the quality of control groups increased across experiments; thus, the differences favoring delayed review were the most difficult to obtain. In Experiments 1 and 1a, nonrepeated items were presented one time for 8 s (or 4 s for some items in Experiment 1a); in Experiment 2, nonreviewed items received multiple study-only repetitions in a self-paced fashion for a limited amount of time; and, in Experiments 3 and 3a, nonreviewed items were presented an unlimited amount of times using test-study repetitions until
each item was successfully recalled. Moreover, in Experiment 3, not only was the nonreview control group given test-study repetitions, but the massed spacing group was given both test-study repetitions and immediate reviews of information. Delayed review was found to be superior to both test-study repetition and immediate review. Thus, for the long-term retention of information, delayed test-study review appears to be the most effective method of restudy.

The results of the present experiments suggest that test-study review is a very powerful learning method that should be employed more regularly within the classroom. Future research will need to determine just how powerful test-study review is and to demonstrate its usefulness within an educational setting. The power of test-study review could be tested by investigating the massive learning of information, such as one-hundred, two-hundred, or five-hundred words, across retention intervals of educational interest, such as one week, one month, or one year. If high levels of recall are found for massive amounts of words, then it would seem within reach to programmatically learn a very large body of information, such as all the words in the dictionary (see also D'Anna, Zechmeister, & Hall, 1991; Zechmeister, Chronis, Cull, D’Anna, & Healy, in press).
REFERENCES


Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with


important lexicon. *Journal of Reading Behavior.*


Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don’t. *Bulletin of the Psychonomic Society, 15,* 41-44.
William L. Cull
Department of Psychology
Loyola University Chicago
Chicago, Il. 60626

Birthdate: 8/15/67
Marital Status: Single
Home: (708) 475-4847

Education

B.S. 1989 Loyola University, Chicago, Illinois
Major: Psychology

M.A. 1992 Loyola University, Chicago, Illinois
Thesis: Metamemory and Learning Ability

Professional Experience

1989-92 Graduate Assistant in Experimental Psychology
1992-93 Part-time Instructor of Psychology
1993-94 Teaching Fellow

Publications


Conference Presentations

meeting of the Midwestern Psychological Association.


The dissertation submitted by William L. Cull has been read and approved by the following committee:

Dr. Eugene Zechmeister, Director
Professor, Psychology
Loyola University Chicago

Dr. Lois Leidahl
Associate Professor, Psychology
Loyola University Chicago

Dr. Emil Posavac
Professor, Psychology
Loyola University Chicago

Dr. Patricia Tenpenny
Assistant Professor, Psychology
Loyola University Chicago

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Ph.D.

12-2-94
Date

Director's Signature