DOCUMENT RESUME

ED 085 420 TM 003 365

| | |
|---|---|
| AUTHOR | Lovelace, Eugene A. |
| TITLE | Effects of Anticipated Form of Testing on Learning. Final Report. |
| INSTITUTION | Virginia Univ., Charlottesville. |
| SPONS AGENCY | National Center for Educational Research and Development (DHEW/OE), Washington, D.C. |
| BUREAU NO | BR-2-C-019 |
| PUB DATE | Aug 73 |
| GRANT | OEG-3-72-0033 |
| NOTE | 19p. |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Essay Tests; *Expectation; *Learning Processes; Measurement Techniques; Objective Tests; *Recall (Psychological); *Recognition; *Testing |

ABSTRACT

This report deals with the effects of an individual's expectations regarding how he will be tested on what he does during learning and what gets stored in memory. It is maintained that essay exams requiring recall are preferable to objective (recognition) tests. There are some bits of empirical evidence as well as some theoretical reasons to believe that recognition and recall memory processes are different; this difference is not only in terms of performance level or mastery of the material which they require, but in terms of what the individual must do to optionally prepare for these two types of tests. A series of nine experiments were conducted; data from this series suggest that in some cases there is only a slight superiority of recall for individual's anticipating the recall task over those expecting a recognition test of memory. (Author/NE)

Final Report





Project No. 2-C-019
Grant No. OEG-3-72-0033

Eugene A. Lovelace
Department of Psychology
University of Virginia
Charlottesville, Virginia 22901



# EFFECTS OF ANTICIPATED FORM OF TESTING ON LEARNING

August 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

Final Report

Project No. 2-C-019
Grant No. OEG-3-72-0033

# Effects of Anticipated Form of Testing on Learning

Eugene A. Lovelace

University of Virginia

Charlottesville, Virginia

August 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

## INTRODUCTION

The present research is concerned with the effects of an individual's expectations regarding how he will be tested on what he does during learning and what gets stored in memory.

The widespread availability of machines for scoring examinations (e.g., the use of IBM sheets) and the frequently high ratios of pupils to faculty in American classrooms have led to an increasing use of various objective tests to measure the student's learning. These tests typically take the form of true-false questions, multiple choice items, or matching exercises. All of these are, in some sense, tests of recognition memory rather than recallability of learned material. It is often assumed that recognition and recall are quite different processes, and that students will prepare differently for recognition and recall tests. Surely anyone who has had to answer students' queries about the nature of the exams in a course can vouch for the fact that students feel that they will prepare differently for different kinds of exams. There is evidence that students report preparing differently for various types of tests (e.g., Terry, 1933; Silvey, 1951).

It is typically maintained that essay exams requiring recall are preferable to objective (recognition) tests, since they lead the students to a greater mastery of the content (e.g., Adams, 1965; Stanley, 1964). As Hakstian (1971) has recently noted, however, this notion is "based on intuitive appeal, but not convincingly supported by empirical research (p.324)."

There are some bits of empirical evidence as well as some theoretical reasons to believe that recognition and recall memory processes are different; this difference is not only in terms of performance level or mastery of the material which they require, but in terms of what the individual must do to optimally prepare for these two types of tests. According to Kintsch (1970), one of the more prominent two-process theorists in this regard, recall contains an active retrieval of items from memory store which is not necessary for recognition tasks. In terms of a distinction maintained by Tulving and his associates (e.g., Tulving and Pearlstone, 1966), any event which has representation in memory is available in memory store, but only those events which the individual can now retrieve are accessible in memory. Obviously an item cannot be accessible unless it is available, but not all materials available in memory are readily accessible. The markedly superior performance with recognition tests, as compared to recall, are generally attributed to the fact that in recognition the accessibility is assured, i.e. the item itself provides the optimal possible cue to gain access to its representation in memory. Therefore recognition is viewed as essentially a measure of what is available in memory, whereas performance on recall tasks requires both availability and accessibility of items in memory. As

Kintsch (1970) has stated: "In recognition...no need exists to consider relationships between the items being learned. Recall learning is quite different in this respect: relationships among items are all-important in recall. The characteristics of a list as a whole rather than the characteristics of individual items determine recall performance. Recall involves a search and retrieval process, the efficiency of which depends upon how well the learning material has been organized in memory (p. 243)."

The addition of the retrieval component has implications for optimal strategies for storage of materials which the individual must recall from memory. In the recall test it is desirable for any items which the individual can retrieve to serve as effective cues to gain access to additional items in memory. That is, inter-item associations of some sort should markedly enhance recall, but not necessarily recognition. In fact, recognition memory may be as good or better under an incidental learning condition than when the individuals expect to be tested for memory of the words (Eagle & Leiter, 1964).

There is evidence, from tasks in which individuals are presented with a list of words in a paced fashion and then asked to recognize the items in a large pool or recall as many as they can, that inter-item associations, or any sort of organization of the words, will facilitate recall but have relatively little effect on recognition of the items (e.g., Cofer, 1967; Kintsch, 1968). That such associative relationships or organizations of items is necessary to recall is intuitively appealing, and agrees well with students' observations regarding the need to organize materials better for an essay test than for recognition tests. The present research is concerned with what an individual does during the learning of a set of verbal materials, and whether this is influenced by the sort of memory test which he expects.

Despite intuitive and theoretical reasons to expect people to attempt to organize or inter-relate materials more when expecting a recall test than when expecting a recognition test, data from a recent study by Hakstian (1971) suggest that no such differences are obtained. However, data from a pilot study in our laboratory using a free recall task clearly suggested that the S's processing of a list of words was influenced by the expected form of testing. For 30-word lists recall performance for Ss set to expect a recognition test was poorer than for Ss expecting the recall task (20.4% vs. 36.5%, $t$ (34) = 5.17, $p$ <.001).

A series of nine experiments were conducted; these entailed free recall for list of words presented either visually or aurally in succession or simultaneously, a paired-associate task involving word pairs, and recall of facts from a prose passage. These experiments seem to confirm the replicability of this pilot data, but also support the suggestion of Hakstian that the expected form of testing is of minimal importance when Ss are learning prose passages. Data from this series of experiments suggest that whenever the study task is presented slowly (or S-paced) and readily permits inter-relating the materials, there is only a very slight superiority of recall for individual's anticipating the recall task over those expecting a recognition test of memory.

## METHODS AND RESULTS

The rationale, procedures and results will be presented separately for each of the nine experiments. Experiments I through VII studied the free recall of common English words under various task conditions; Experiments VIII and IX involve paired-associate learning of word pairs and recall of materials from a prose passage, respectively.

The Ss in all experiments were drawn from introductory psychology courses and were typically run in small groups of 2 to 5 per session. They received course credit for participation.

### Experiment I

In this initial experiment the S's expectations ("set") regarding the form of testing were determined both by instructions and by the preceding task given in the laboratory. Half the Ss were set to expect recall and half recognition; for each of these groups half received a recall test and half received a recognition test.

Method.

Materials. The materials employed were 180 nouns taken from the norms of Paivio, Yuille and Madigan (1968). All words had a frequency greater than 20 with imagery, concreteness and meaningfulness above 2.5, 2.9, and 4.0 respectively. The 180 words were divided into 3 base lists of 60 words each; the words were chosen so as to eliminate obvious associations among words within each list, and between lists. All words used contained between 5 and 9 letters.

Design and Procedure. Each of the three lists was subdivided into an A and a B portion with 30 words in each. The Ss were randomly assigned to either the A or B form upon entering the laboratory. Those Ss assigned to a recognition condition received words from the other form as foils during the recognition phase. For example, Ss receiving form A words as the study list received words from form B as distractors during recognition. All Ss received the words of each list in the same order, with words presented at a 2-sec. rate. (The complete set of materials are available from the author).

The Ss were randomly assigned to one of four conditions (N=20/condition). Each S was presented with three lists of words regardless of the condition to which he was assigned and he was tested for retention after each list was presented.

Condition R1-Rn: Before presentation of the first list, Ss in this group were informed that their task would be to recall as many of the words presented to them as possible. The 30 words were then presented one at a time. After presentation of the list, Ss were instructed to write down in any order, the words they remembered. They were given 5 min. to do this and then were informed that a second list would follow. Once again Ss were instructed to expect a recall task, and after presentation of the 2nd list they were tested for retention. Before the final list, the instructions given to the Ss implied, but did not state, that they would

be asked to recall after the third list. Following list 3 Ss were given a recognition booklet containing 30 word pairs, one pair per page; the ordering of these pages was varied over Ss. The Ss were instructed to circle the word from each pair which had been presented during the third study list. (Verbation instructions for all conditions are available from the author).

Condition Rn-R1: Before list 1, Ss in this group were informed that their task would be to later recognize which word in a given pair belonged to the list of words they would study. After the list was presented, Ss were given recognition booklets and asked to select the word in each pair which had been presented. Following this recognition test, Ss were informed that they would receive a second list and would again be required to recognize the presented words from a pair. The test procedure following list 2 was the same as that following list 1. The Ss then were instructed to prepare for a third and final list which implied that the test form would again be a recognition booklet. Following this third list, however, Ss were given a blank sheet of paper and asked to recall as many words as possible. They were allowed 5 min. to recall and told they would not be penalized for incorrect answers.

Condition Rn-Rn: This group served as a control for the R1-Rn group and received a recognition test on all three lists. The Ss in this group were informed before the first list that their task would be to recognize the test words from a given pair. After each of the three lists they received recognition booklets in which they circled the correct words.

Condition R1-R1: This group served as a control for the Rn-R1 group. The Ss in this condition were set to expect, and did in fact receive, the recall test described above on each of the three trials.

Apparatus. All words were projected by a Kodak Carousel 800 projector onto a wall-screen; rate of presentation was controlled by a Lafayette Model 4B repeat-cycle timer. Instructions for each stage of the experiment were presented on a cassette tape recorder.

Results.

Mean number of correct recall and recognition responses for each condition on each of the three trials are shown in Figure 1. Groups which received recall or recognition on all trials showed no appreciable change over trials. Recall performance did not differ between the R1-Rn and the R1-R1 conditions on trials 1 and 2, nor did recognition performance differ between the Rn-R1 and the Rn-Rn conditions over these first two trials.

The major interest, of course, centers on performance on the third trial. A comparison of Rn-R1 and R1-R1 groups showed that those Ss expecting recall retained about 25% more words on trial 3 than those expecting recognition (13.10 vs.9.75). This difference was statistically significant, $t(38) = 2.33$, $p < .05$. A comparison of recognition performance on the third trial indicated that the Rn-Rn group did not differ from the R1-Rn group. (Due to the skewed, non-normal nature of

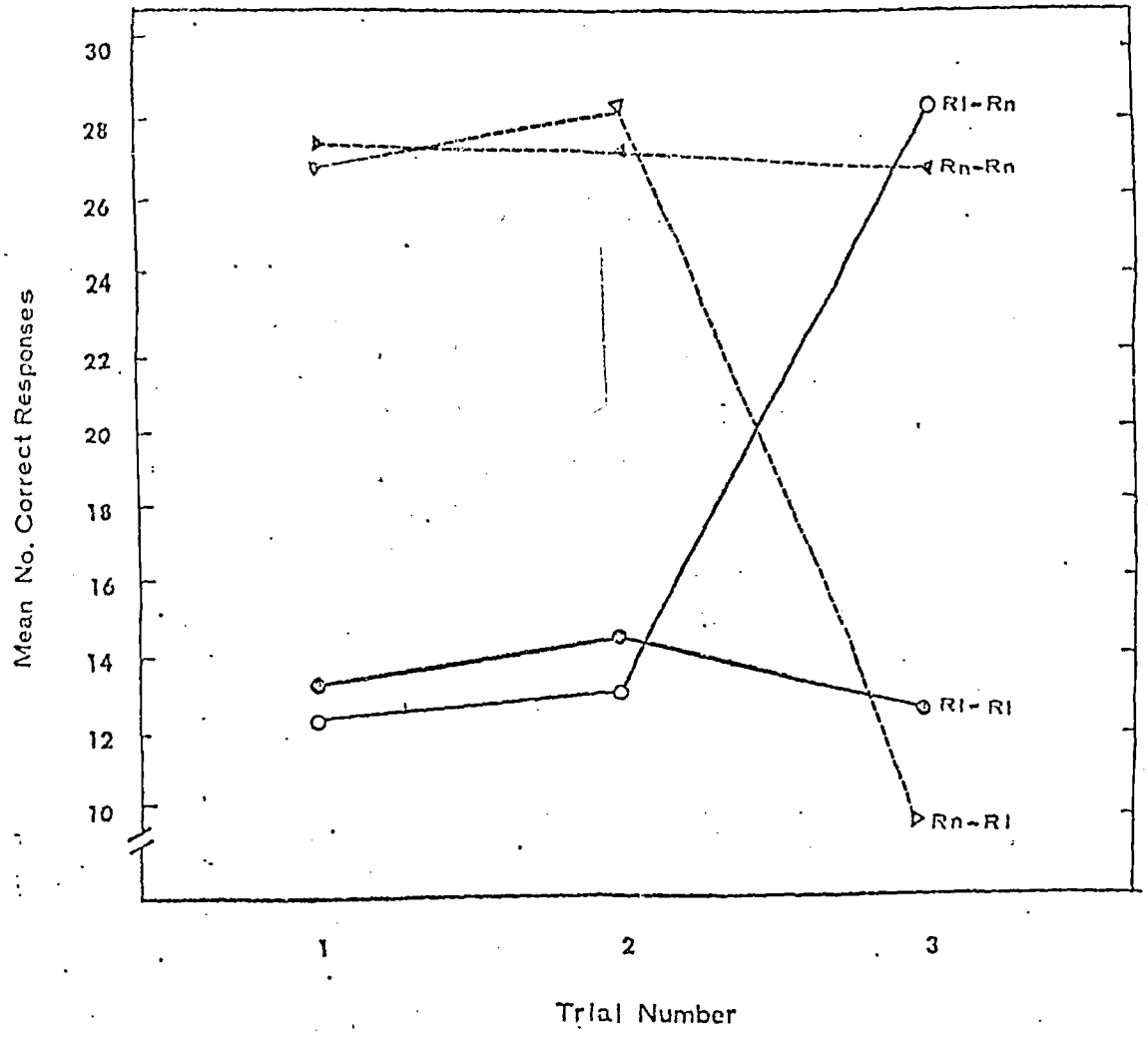


Fig. 1. Mean numbers of correct recall and recognition responses for Experiment I for all four conditions.

this distribution, where nearly half the Ss made no errors, recognition scores were compared by a Mann Whitney U test, U = 167, $p > .1$).

The mean percentage of Ss correctly recalling words as a function of the input positions of those words is shown in Figure 2. Input positions include the words occupying that position for both the A and B form, and are collapsed over blocks of three adjacent words. The significant superiority for Ss expecting recall appears to come primarily from primacy and recency positions, i.e., from the first few and last few words in the list. When the last five items were eliminated from the comparison, the difference between the R1-R1 groups was no longer significant, $t$ (38) = 1.69, $p$ = .10. Since all Ss received the words in the same order, input positions were perfectly confounded with specific words. Thus, the sizable difference in performance on items from late in the list might be a materials effect or a recency effect, in the sense of differential use of active memory as a basis of recall. If it were the latter it should show up in the output order during recall; that is, the words from these recency positions should be "spewed" as initial items in output by Ss in group R1-R1, but not by Ss in group Rn-R1. An examination of output orders gave no evidence of such spewing. This suggests that the superiority of R1-R1 to Rn-R1 for those items late in the list more likely reflects properties of the items themselves rather than the input positions per se.

Intrusion errors calculated on trial 3 for the Rn-R1 and R1-R1 groups indicated that expectancy had little effect on the occurrence of such errors in recall. The mean number of intrusion errors was 1.15 for the R1-R1 group and 1.55 for the Rn-R1 group. Approximately one third of the Ss in each group made no intrusion errors.

## Experiment II

The data from Experiment I indicate differences in recall performance as a function of S's expectations, and a suggestion that this is primarily due to the "recency" positions. Experiment II was a replication of Conditions Rn-R1 and R1-R1 with presentation order of the words counterbalanced across Ss so as to eliminate the confounding of specific words with input positions.

### Method.

Materials. The materials employed were the same three lists used in Experiment I. Four different presentation orders were used for the third list; these orders were derived in the following manner. The last five words in the original list were distributed in a random manner within the other words in the list and new words were placed in the terminal five positions. This ordering made up the first transformation. For the second transformation, the last five words were again redistributed among the other 25 and five new words were chosen to occupy the terminal positions. This procedure was carried out until four transformations of the original list were formed such that, in comparing the four lists, no word appeared in the last five positions more than once and the remaining words were unsystematically re-arranged within the list for each transformation.

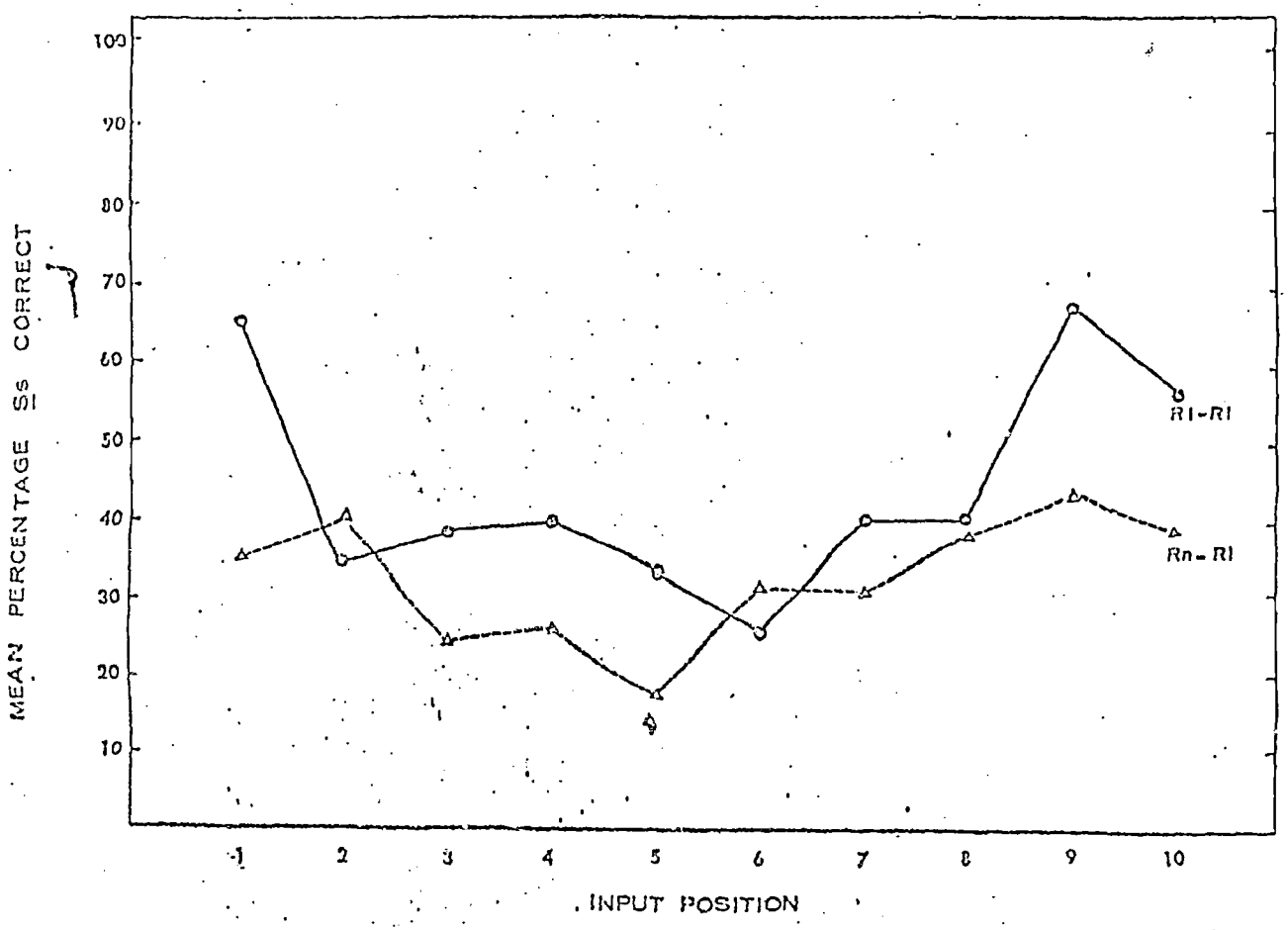


Fig. 2. Percentage of Ss in Experiment I recalling words as a function of input position. The data are collapsed such that each position represents three items of the study list.

Procedure and Apparatus. The procedure and apparatus for both groups (Rn-Rl and Rl-Rl) were the same as for their respective groups in Experiment I. Thirty-two Ss were randomly assigned to each condition.

Results.

A comparison of recall scores on the third trial revealed a difference of approximately 18% in the expected direction, with Ss expecting recall performing better than those expecting recognition (10.97 vs. 9.06). This difference only approached the conventional level of statistical significance, however, $t(62) = 1.8$, $p < .1$. Inspection of Figure 3 indicates that the two groups do not differ most in the output of the terminal items in the list, and primacy effects are apparent in the two conditions. Whatever differences do exist between the Rn-Rl and Rl-Rl groups is apparently not attributable to differential recall of the last items in the list.

Analysis of the intrusion errors produced by the two groups reaffirmed the similarity of their performance despite their examination set. About half of the Ss in each group gave no intrusions, with the mean number of such errors being 1.25 for Ss in group Rl-Rl and 1.41 for Ss in group Rn-Rl.

## Experiment III

The first two experiments substantiate the finding in the pilot study that Ss who are expecting a recall test can free recall more words than those expecting a recognition test. However, the design of those experiments is such that Ss in the Rn-Rl and Rl-Rl conditions also have differential practice with the recall task in the experimental situation. Although college students have undoubtedly had a great deal of practice with recall tests, and the stable performance of Rl-Rl across the three lists gives no evidence of any "learning-to-learn" phenomenon, an experiment was designed to eliminate this confounding.

In Experiment III Ss received a single list with the expectation regarding the form of testing being induced solely by instruction. Three other changes were made: a) the study list was longer, composed of 60 rather than 30 words, b) each word was presented for 3 sec. rather than 2 sec. and c) a numerical task was interposed between study and recall. If organizational factors are important to the level of recall performance, permitting Ss to organize material is probably crucial to the superior performance found when Ss are expecting a recall test. The 2-sec. rate may not have permitted Ss sufficient time to optimally organize the material. Thus, the differences in performance may have been minimized by not allowing time for meaningful reorganization of the word list. The limited amount of time would not have such an adverse effect on the performance of Ss expecting recognition, if they typically do not make much use of organizational processes in learning.

Lengthening of the list from 30 to 60 words should also reduce the "ceiling" effects in recognition performance which posed an interpretational problem for recognition data from Experiment I.

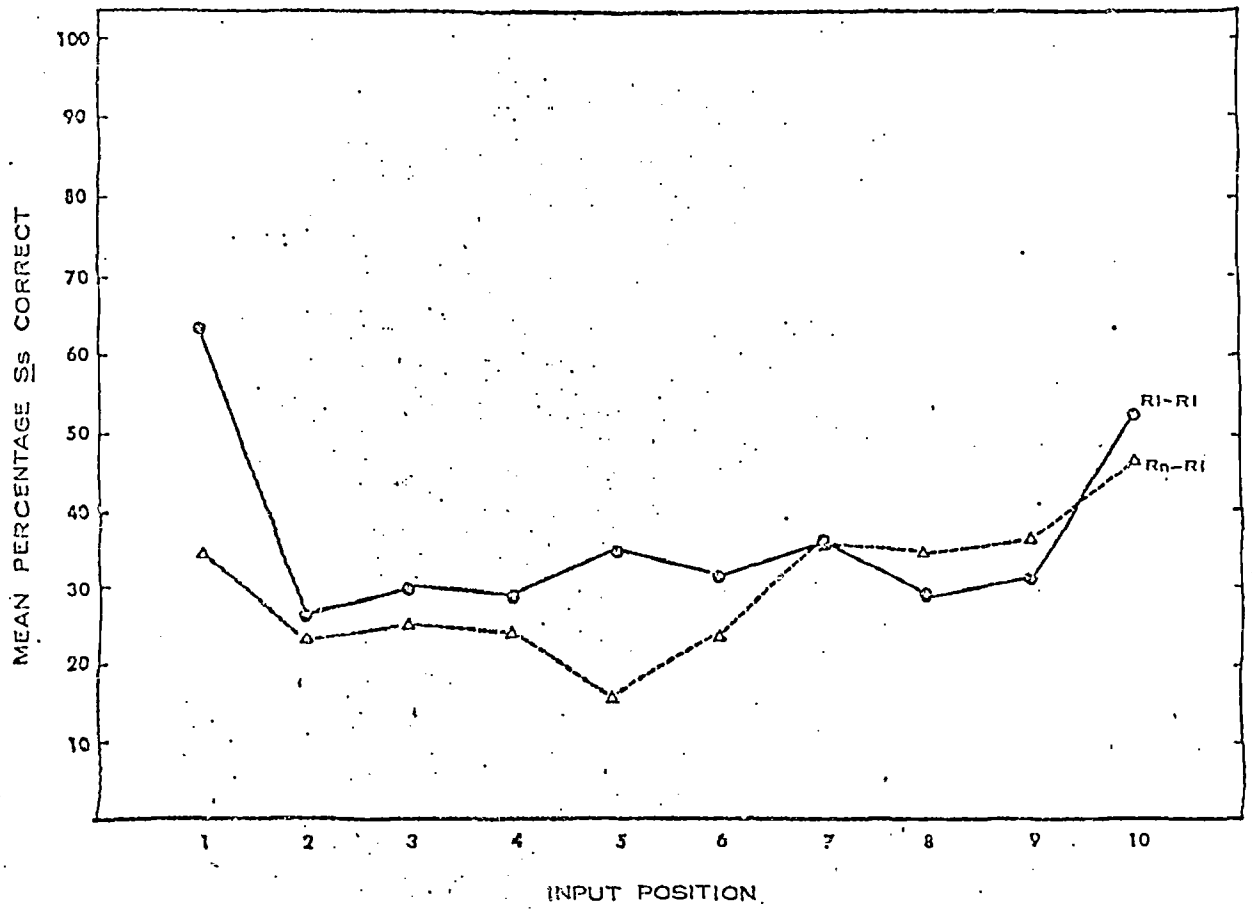


Fig. 3. Percentages of Ss in Experiment II correctly recalling words as a function of input position. The data are collapsed such that each position represents three items of the study list.

Method.

Materials and Design. The materials used were chosen from the 180 words of Experiment I. Only one 60-word study list was employed, composed of the combined A and B forms of list 3. For groups receiving a recall test, four transformations of word order were used to minimize any effects of presentation sequence. The first and last 7 words in the list were redistributed for each transformation such that no word occupied either of these list portions for more than one quarter of the Ss. The remaining words were randomly ordered throughout the other 46 positions. For Ss tested by recognition a single word order was employed. Distractor items for the recognition pairs were formed by combining the A and B form of list 2 and using these 60 words as foils for the test list. A mathematical task consisting of approximately 100 addition and subtraction problems was constructed for Experiment III. The problems, composed of two or three 5-digit numbers, were introduced following the study list to prevent Ss from using the time before testing to rehearse the items which had been presented; this should assure that performance was not based on active short-term memory.

Procedure. The 80 Ss were randomly assigned to one of the following four conditions, N = 20/condition.

R1-Rn: Before presentation of the list, Ss in this group were instructed to prepare for a recall task. The words were presented at a 3-sec. rate, after which Ss were given 3 pages of mathematical problems to compute. They were allowed 3 min. to work on this task; there were about 30 problems per page, more than any S was able to complete. Recognition booklets were then given out. The Ss were allowed as much time as needed to circle the correct word of each pair in the booklets, and then were given a sheet of paper on which to recall as many words in the list as possible.

Rn-Rn: Procedure for this group was the same as for the R1-Rn group except for their initial instructions. Before presentation of the list, Ss were told to prepare for a recognition test.

Rn-R1: Initial instructions for this group indicated that their task would be to recognize words in the test list when paired with distractor items. After the 60-word list was presented, and the mathematical task performed, Ss were instructed to try to recall the words they had seen in the list. They were given 4 min. in which to do this, and then were given recognition booklets to complete as their own pace.

R1-R1: The Ss in this group were instructed to prepare for a recall test. After seeing the list and performing the mathematical task, they were permitted 4 min. for recall. Following this recall, the Ss were allowed to work through the recognition booklet at their own pace.

The equipment used was the same as that used in Experiments I and II.

Results.

Recall performance for the R1-R1 and the Rn-R1 groups proved to be

significantly different, $t(38) = 2.57$, $p < .02$. Comparison shows this difference in mean correct responses to be about 28% in the expected direction (15.85 vs. 11.50). This difference is of slightly greater magnitude than that obtained in Experiment I (28% vs. 25%) and did not result from differences at a few particular input positions. Thus the differences in recall observed in the first two experiments were properly attributed to the expected form of the test and not to strategies developed across the three lists: differences in the present experiment can not be seen as a learning-to-learn phenomenon as all groups received only one list and thus differed only with respect to their anticipation of test form.

A comparison of recognition performance for Rl-Rn, Rn-Rn groups revealed that recall-set Ss recognized more items then recognition-set Ss (54.70 vs. 53.97), but this small difference did not approach significance, $t(38) = .46$, $p > .5$.

It should be noted that a second test was given all Ss after the primary manipulation of the experiment took place. This second test was introduced solely to fulfill the instructional set the Ss in two of these four groups (i.e., Rn-Rl and Rl-Rn) received prior to testing; no further consideration will be given here to those data.

## Experiment IV

The replicability of the superior free recall of words when Ss were expecting a recall test to that when expecting recognition seems clearly established by Experiments I - III. In order to attempt a direct assessment of any organization which the S is imposing during the study period, it was decided to present the words auditorally and ask Ss to write these down for later study. (See Experiment V for rational, procedure, etc.) Before undertaking such a study, however, it was necessary to establish that the effects of anticipated form of test which were found in the first three experiments were not modality-specific. Experiment IV provides a replication of Experiment III with the words auditorally presented.

### Method.

The materials, design and procedure were exactly as in Experiment III except that the words were presented auditorally from a tape recorder instead of being projected on a screen, and the presentation rate was slowed from 3 sec./word to 4 sec./word.

### Results.

The mean number of correct responses for each of the four conditions were: Rn-Rn = 53.6; Rl-Rn = 54.8; Rl-Rl = 17.7; Rn-Rl = 13.4. The effect on recall performance of the expected form of test (17.7 vs. 13.4) is a 24% difference; this is very close to the values obtained in Experiments I and III. Due to a slight increase in variance, however, this difference does not quite attain the conventional level of statistical significance, $t(38) = 1.91$, $.05 < p < .10$. The comparability of these results to those of Experiment III, however, lead to the conclusion that this phenomenon is not modality-specific.

The very small difference in recognition performance as a function of expected form of testing also replicates the results for visually-presented word lists. Although recognition performance here is good (again around 90%) it is doubtful that the absence of a difference here is an artifact of a performance ceiling. Rather it appears that for words presented either visually or auditorally the recognition performance is not substantially related to expected form of the test. (It should be pointed out, perhaps, that in Experiments I, III and IV the Ss expecting recall actually performed slightly better than those expecting recognition--the direction of any effect has been consistent, but the magnitude of the effect very small.)

## Experiment V

If the superior recall of a word list by Ss expecting recall rather than recognition is due to some greater degree of organization of the to-be-remembered items when a recall task is anticipated, it might be possible to assess the difference in organization which Ss of the two groups impose during the study period. The present experiment was an effort to do that by having Ss write down the words (which were auditorally presented) for later study. It was hypothesized that Ss expecting recall would not simply record the words in order, but that their associative organization would be reflected in the spatial array of the words as written on the paper, i. e., Ss expecting recall would write down "related" words in adjacent places on the paper whereas Ss set for a recognition test would simply write the words of the list in the order in which they were presented.

### Method.

The materials, apparatus and design essentially replicate the Rn-Rl and Rl-Rl conditions of Experiment IV but with the following changes. The Ss were given a blank sheet of paper prior to the study phase and instructed to write down each word, as they heard it spoken. They were told that the words "need not be listed in the same order in which they are presented, as different individuals are receiving the words in various random orders.....After you have heard all 60 words you will be given 1 min. to look over the complete written list of words which you heard." Following this 1-min. study of their written copy of the list the Ss were given a blank sheet of paper and asked to write down all the words which they could recall.

In order that the Ss have ample time to find on their page any other items to which a presented word seemed "related", the presentation was at 8 sec./word. (A pilot study at 4 sec./word with six Ss in each condition showed no differential organization, and Ss appeared to be in somewhat of a rush to get words written down at that rate. Such concern about keeping up with the task would clearly preclude the words being recorded in ways that might reveal any organization S was imposing on the list, thus the rate was slowed to 8 sec./word.)

Fifteen Ss served in each condition.

Results.

The results of this experiment are most succinctly summarized as "not very informative." The study sought evidence of differential organization occurring for Ss who recall differentially due to expectations about the form of testing. There was no evidence of differential organization, but neither was R1-R1 recall superior to Rn-R1, $\bar{x}_s$ = 24.26 and 23.73, respectively, $t$ (28) < 1. The combination of slower presentation and the S actively recording each word, plus the 1 min. of review, led to higher performance levels, but with these increased opportunities for study and organization the benefits of anticipating a recall test were essentially eliminated.

Given the absence of differences in recall, little difference in organization would be expected. However, the comparability of the two groups' performance in recording the words they heard did not derive from equal evidence of organization. There was essentially no evidence of organization with either test set. The Ss of both groups simply recorded the words in order as they heard them. This technique for assessing differences in organization is not only insensitive to organization (none was apparent in the protocols) but may actually have eliminated the phenomenon it was designed to assess.

Experiments VI and VII

One of the ways in which Experiment V differs procedurally from the earlier studies is that, by time of recall, the S has had an opportunity to study the list with all items simultaneously present. This may change the S's strategy from that employed when items are presented for study singly and successively; even though in simultaneous presentation the S still must successively read the words, the opportunity for selective review, and so imposition of organization, seems greater. Experiments VI and VII essentially replicate the Rn-R1 and R1-R1 conditions of Experiment III except that in both of these experiments (VI & VII) the list of words to be free recalled was presented simultaneously rather than successively (hereafter referred to as "whole-list" presentation).

Method.

Apparatus and Materials. The list of words to be recalled in Experiment VI was the same list as was used in Experiment III (i.e., list 3 of Experiment I), whereas the list to be recalled in Experiment VII was list 2 of Experiment I. Instead of the materials being projected on a screen for study the S studied the words from a single sheet of 8½ x 11 in. paper on which the words appeared in three columns, 20 words per column. The words were typed, with only initial letters capitalized, with triple spacing between words of a column and about 2 in. separating the columns.

Procedure. The Ss were given tape-recorded instructions to induce either a recognition (Rn-R1) or recall (R1-R1) test set, and then allowed 4 min. to study the list of words (this is equivalent, in terms of total study time, to 4 sec./word). The Ss performed a math task for 3 min. following the study interval to minimize short-term memory difference;

they were then allowed 4 min. to write on a blank sheet as many of the words as they could recall.

Results.

The mean number of correct responses in the Rn-R1 and R1-R1 conditions for Experiment VI were 15.3 and 17.5, respectively. For Experiment VII these values were 21.3 and 22.8, respectively. Although performance was better for the words employed in Experiment VII than that of Experiment VI the direction and magnitude of the differences as a function of test set are very comparable. In the absence of any evidence of an interaction, the data of the two experiments were combined to assess the overall effect of test set for recall of a word list using a whole-list presentation procedure. The combined mean correct responses were 18.6 and 20.5 for the Rn-R1 and R1-R1 conditions; this difference does not approach statistical significance, $t$ (38) = .81.

## Experiment VIII

Although there is evidence that the learner's test set can, under some conditions, influence the amount which that individual may be able to free recall, such findings may be of limited generality. Is the phenomenon restricted to the free recall task or would it also occur for a task which has an explicit task requirement of associative learning? It was the purpose of Experiment VIII to examine the effects of anticipated form of test for verbal materials involved in a paired-associate task.

Method.

Apparatus and Materials. The 60 words of list 2 in Experiment I served as stimulus members of a 60-pair list while the 60 words of list 3 served as response terms; pairing was by random assignment. The word pairs were typed and photographed with the stimulus word above the response word and projected on a wall screen for Ss to study at a 4-sec. rate.

Procedure. Twenty Ss were instructed so as to induce a recall set, and 20 Ss were given a recognition set. In R1-R1 the Ss were told they would later be shown the top word of each pair and have to write down the word shown with it; in Rn-R1 they were told they would be tested by being given a booklet with the top word of each pair printed to the left and three choices printed to the right including the word it had been paired with and they were to circle that word. Both groups were tested upon completion of the study trial by being shown each stimulus word alone on the screen and asked to recall and write down the word which went with it in successive blanks of a test sheet. These stimulus words were in a different random order than in the study trial; this recall test was paced at a 4-sec. rate. Following this recall test the Ss in Rn-R1 condition were given the test booklet originally described, just to maintain the integrity of the E's original instructions to this group.

Results.

The mean numbers of response words correctly recalled was 12.20 for Ss in the R1-R1 group and 11.25 for Ss in the Rn-R1 group. This small difference yields a $t$ of less than unity; there is no apparent effect

on recall performance in a paired-associate task of the Ss expectation of recall vs. recognition tests.

## Experiment IX

It might be argued that the tasks of Experiment VIII was more like "real-world" learning since there were explicit, new associative connections to be formed. In a great many respects, however, the paired-associate task is as artificial in its tasks characteristics as is free recall. Most notably, although all of the experiments in this series have used verbal materials, none had those materials presented in a prose context. Thus the to-be-learned material never occurred with the usual contextual, semantic and syntactic richness which typifies most verbal materials the individual might study. In Experiment IX the Ss studied a prose passage with either a recognition or a recall test set, and then attempted to recall answers to a series of short-answer, fill-in-the-blank questions.

Method.

Materials. The Ss studied a 12-page, 3000 word passage from a popular book concerning aquatic life; this passage was double-spaced on 8½ x 11 in. paper. The recall test was composed of 25 fill-in-the-blank completion items. These test items were sentences verbatim from the prose passage, or close paraphrases of these sentences, with the critical fact left blank for the S to write in.

Procedure. Eighteen Ss were randomly assigned to each of two conditions. The Ss received a booklet with instructions on the first page about the study task and about subsequent testing procedures. For one condition these instructions set the Ss to expect recall--"a number of short answer - fill in questions about the passage." The Ss of the other condition were led to expect recognition--"a number of multiple choice questions about the passage." The instructions also indicated that the Ss were to spend as much time reading the passage as they felt necessary, but that a maximum of 30 min. would be allowed. Each S was asked to record the time at which he started reading the passage and the time at which he finished reading and went on to the test. All Ss were given the recall test immediately upon completion of studying the passage; Ss given the recognition set were given a multiple-choice test on the same questions following the recall.

Results.

It was anticipated that there would be differences in reading times as a function of test set, i.e., that Ss expecting a recall test would spend more time studying the passage than would Ss expecting a recognition test. This could, of course, lead to interpretational difficulties if Ss expecting recall did in fact recall more items than those expecting recognition. The mean number of minutes spent studying the 12-page passage was 16.7 for Ss expecting recall and 17.0 for those expecting a recognition test. The distributions of reading times were nearly identical for the two conditions.

The mean number of correct respones on the short-answer completion

test was 11.39 for Ss expecting recall and 11.22 for those expecting recognition. Test expectation, as induced by an instructional set for fill-in vs. multiple choice test, clearly had no detectable effect on either reading time or the amount the Ss could recall.

## DISCUSSION AND CONCLUSIONS

Evidence existing before these studies were conducted, including the work of Hakstian (1971), makes it clear that students prepare differently for course examinations when they expect a recall test than when they expect of recognition test. The introspective reports of the great major of our Ss support this notion; when the purposes of the present research were described during de-briefing at the end of each session most of these Ss agreed that they prepare differently for the two types of tests. The results of the present experiments, however, indicate that there are limitations on the conditions under which memory test performance itself will be influenced by these differential expectations.

It would appear that when the to-be-learned material contains very little intrinsic organization the Ss do much better if they expect a recall test than when they expect a recognition test. Thus in the free recall experiments, regardless of input modality there was superior performance for Ss expecting recall than for those expecting recognition. These free recall lists were composed of words which the Es judged to be associatively "unrelated"; for such material it is presumably important for the S to try to impose an organizational schema which will facilitate the retrieval of these otherwise un-associated items, and expectation of a recall test may result in such organization.

When the nature of the task is such as to assure that all the Ss do attempt to form associative connections, then the advantage for recall of a recall set, rather then a recognition set, is lost. In the paired-associates task, even though the same pool of "unrelated" words were used as in the free recall task, the imposed requirement that Ss attempt to associate these words in pairs led to essentially an elimination of effects of anticipated form of testing.

The presentation of a prose passage as the to-be-learned material also seems to minimize effects of test expectations. This finding is consistent with the recent report of null results of this manipulation by Hakstian (1971) and suggests that although the nature of the anticipated form of testing may substantially change the way in which an individual goes about studying the to-be-learned material, it may not have any very noticeable impact on how much he has learned as reflected in short-answer recall sorts of tests. One must be cautious, however, about over-generalizing. The present results should not be taken simply as evidence that it doesn't matter what sort of test the individual expects. A question can be raised as to whether the dependent measure may have been insensitive to differences in what the Ss were able to learn about the passage. The present null result might obtain for short-answer recall, but not for an essay test. That is, if the memory test was one which provided the S with

fewer specific retrieval cues, and thus forced the individual to rely more heavily on an overall organization of the material, a recall expectation might have produced superior performance. This possibility remains to be tested with our prose materials.

## References


Adams, G. S. Measurement and Evaluation in Education, Psychology, and Guidance. New York: Holt, Rinehart and Winston, 1965.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. J. Kleinmuntz (Ed.) Concepts and the Structure of Memory, New York: John Wiley, 1967.

Eagle, M., & Leiter, E. Recall and recognition in intentional and incidental learning. Journal of Experimental Psychology, 1964, 68, 58-63.

Hakstian, A. R. The effects of type of examination anticipated on test preparation and performance. The Journal of Educational Research, 1971, 64, 319-324.

Kintsch, W. Recognition and free recall of organized lists. Journal of Experimental Psychology, 1968, 78, 481-487.

Kintsch, W. Learning, Memory, and Conceptual Processes. New York: John Wiley, 1970.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, Part 2, 1-25.

Silvey, H. M. Student reaction to the objective and essay test. School and Society, 1951, 73, 377-378.

Stanley, J. C. Measurement in Today's Schools (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1964.

Terry, P. W. How students review for objective and essay tests. The Elementary School Journal, 1933, 33, 592-603.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 381-391.