Bulletin of the Psychonomic Society
1978, Vol. 11 (2), 126-128

Evaluation of a method for studying forgetting: Is data from split-half recognition tests contaminated by test interference?

DONALD BAMBER and VICTOR MANIFOLD
Research Service, Veterans Administration Hospital, St. Cloud, Minnesota 56301

Split-half recognition testing is a method of investigating forgetting. In this method, a subject studies a list of items and then is given a recognition test covering half the items on the list. At a later time, a final recognition test covering the other half of the items is given. The present study investigates whether data from split-half recognition tests is contaminated by test interference. Specifically, this study investigates whether taking the first recognition test affects the subject's recognition accuracy on the second test. It is concluded that the first test does not affect performance on the second test provided that the two tests are separated by 24 h or more.

Forgetting is a lessening over time of the material retained in memory. In order to study forgetting, it is necessary to measure retention at two or more different times. This poses a technical problem. The problem is that the earlier retention measurements may interfere with the later measurements. What are needed for the study of forgetting are methods for making mutually noninterfering measurements of retention at various times.

Mutually noninterfering tests of retention are particularly important when investigating individual differences in forgetting. Otherwise, individual differences in susceptibility to test interference may be mistaken for individual differences in forgetting.

Some possible methods for making mutually noninterfering retention measurements will now be considered.

Group-Difference Method

In this method, two or more groups of subjects are randomly drawn from the same population and are given identical learning experiences. Each subject's retention is tested only once, with the different groups being tested at different times following learning.

This method is completely free of any interference between retention measurements made at different times. Unfortunately, it is a between-subjects method of studying forgetting. A within-subjects method would be preferable for two reasons. First, experiments using within-subjects designs generally require fewer subjects than experiments using between-subjects designs. Second, the use of within-subjects designs would facilitate the study of individual differences in forgetting.

This work was supported by the Medical Research Service of the Veterans Administration. The authors are indebted to Rebecca Berg for improving the readability of the paper and to Teresa Kucala for typing the manuscript.

Series-Average Method

This is a commonly used within-subjects method for studying forgetting over brief durations. Each subject is given a series of trials. In each trial, the experimenter randomly selects an item from a pool and presents the item to the subject for study. Then, after a brief interval, the subject's retention of the item is tested. Various retention intervals are employed. Each appears in numerous trials which are evenly distributed throughout the series. For each retention interval, the subject's average retention is calculated.

This method does not eliminate test interference. Rather, it eliminates *differential* test interference; it insures that the amount of interference is approximately equal at all retention intervals.

Unfortunately, this method is not practical for studying forgetting over long durations. Long retention intervals would extend prohibitively the total duration of the series of trials.

Learning-Curve Extrapolation

Another within-subjects method for making mutually noninterfering measurements of retention has been described by Underwood (1964). This method may be used in studies of either paired associate learning or serial learning. A delayed retention test is administered to the subject some time after the completion of learning. The subject's score on the delayed retention test is compared with an *estimate* of what his score would have been on a retention test immediately following learning. Roughly speaking, this estimate of the immediate retention score is obtained by extrapolating the subject's learning curve. The subject's score on the delayed retention test cannot be affected by an immediate retention test since none is given.

Split-Half Retention Testing

Split-half retention testing, still another within-

subjects method, is *possibly* a way of making mutually noninterfering measurements of retention. In this method, a subject studies a list of items and then is given a retention test covering a random selection of half the items. At a later time, he is given a second test covering the other half.

There are several ways in which split-half retention tests may be performed. (1) In split-half tests of paired associate recall, the subject is requested to recall the appropriate responses to half the stimuli on one occasion, and to the other half on a later occasion. (2) Similarly, in split-half tests for recall from a categorized word list, half the categories are specified and the subject is requested to recall the list words from those categories. On a later occasion, the subject is requested to recall the list words from the remaining categories. (3) In split-half tests of recognition, an ordinary recognition test is divided into two parts that the subject takes on different occasions.

Are split-half retention tests mutually noninterfering? This question must be answered empirically. Moreover, a separate answer is required for each type of split-half retention test listed above. The purpose of the experiments reported here is to determine whether split-half tests of *recognition* are, in fact, mutually noninterfering.

METHOD

The two experiments reported here were identical except for the differences noted below.

Subjects

The subjects were male psychiatric patients at the St. Cloud Veterans Administration Hospital. Only patients having no mention of memory problems in their medical records were included in the study. This does not guarantee that the patients in the study had completely normal memories; it does, however, guarantee that their memories were not grossly abnormal. The subjects were all volunteers and received no pay for their participation. There were 44 subjects in Experiment 1 and another 40 in Experiment 2.

Study Lists

A word pool was constructed from adjectives that occurred 20 or more times in the Lorge magazine count of approximately 4.5 million words (Thorndike & Lorge, 1944). Study-list words and distractor words for the recognition tests were drawn randomly from this pool.

The lists were presented to the subjects on a memory drum at the rate of 4 sec per word. They were typed in blocks of five words with a single space between blocks. Spaces relieved the subjects of strain from continuously attending to a long list.

The term "list body" will refer to the entire list exclusive of the first block and the last block. In Experiment 1, the list body consisted of 100 words that were presented once each; another 5 words were presented in the first block and another 5 were presented in the last block. In Experiment 2, the list body consisted of three random permutations of a set of 50 words; another 5 words were presented in the first block and were repeated in the last block.

Recognition Tests

Standard recognition tests were constructed for Experiments 1 and 2. These consisted of 110 and 55 multiple-choice items, respectively. Each multiple-choice item contained one word from the study list and three distractor words. Split-half recognition tests were constructed by randomly dividing the standard recognition tests into two parts. These parts contained equal numbers of list-body words and were labeled Split Half A and Split Half B.

The split-half recognition tests were administered to the subjects in a forced-choice fashion. Because words from the first and last blocks would particularly benefit from any primacy and recency effects, the subjects' recognition responses to these words were not scored.

Procedure

Each subject was paired with another subject having similar age and education. Then, one member of the pair was randomly assigned to the experimental group and the other to the control group. All subjects were tested individually in two sessions 24 h apart. On Day 1, all subjects were shown the study list on the memory drum. Then they worked on addition problems for 2 min. At this point, the control subjects were dismissed, whereas the experimental subjects were given their first split-half recognition test. On Day 2, the experimental subjects were given their second test, whereas the control subjects were given their first test followed immediately by their second test. Half the subjects in each group were tested on Split Half A followed by Split Half B, and half in the opposite order.

RESULTS AND DISCUSSION

Interference Between Split-Half Recognition Tests Separated by 24 h

The critical question is: Would the experimental group have performed differently on the Day 2 split-half test if it had never taken the earlier split-half test on Day 1? To answer this question, the experimental group should be compared with the control group, which took no test on Day 1. Recall that the experimental group took only one split-half test on Day 2, whereas the control group took two. Therefore, the experimental group's Day 2 performance should be compared with the performance of the control group on the *first* of its Day 2 tests.

Table 1 contains the relevant data. Consider the difference between the Day 2 performances of these groups: the experimental group on its single test and the control group on the first of its two tests. This difference went in opposite directions in the two experiments. In Experiment 1, the difference was −.045, with a standard error of .043. In Experiment 2, the difference was +.014, with a standard error of .050. Neither difference was significant [$t(21) = 1.05$, $p > .05$, and $t(19) = .28$, $p > .05$, respectively]. Combining the

Table 1
Proportion of List-Body Words Recognized

	Experiment 1			Experiment 2		
		Day 2			Day 2	
Group	Day 1	Test 1	Test 2	Day 1	Test 1	Test 2
Experimental	.530	.409		.772	.574	
Control		.454	.375		.560	.512

Note—*Chance level = .250.*

difference estimates from the two experiments yielded a joint difference estimate of −.016, with a standard error of .033.

Thus, within the limits of measurement error for this study, split-half recognition tests given 24 h apart appear to be mutually noninterfering.

Interference Between Split-Half Recognition Tests Separated by a Brief Period

Are split-half recognition tests mutually noninterfering if they are separated by a brief period, say, less than 1 min? This question may be answered by comparing the control group's performance on its first split-half test with its performance on the test immediately following. Consider the change in the control group's recognition accuracy between these two tests. This change was negative in both experiments. For Experiment 1, the change was −.079, with a standard error of .025. For Experiment 2, the change was −.048, with a standard error of .037. The change was significant for Experiment 1 [$t(21) = 3.09$, $p < .01$], but not for Experiment 2 [$t(19) = 1.29$, $p > .05$]. Combining the change estimates from the two experiments yielded a joint change estimate of −.064, with a standard error of .022.

It appears that taking one split-half recognition test causes a person to perform poorly on a second test taken immediately after the first. However, if the second test is delayed for 24 h, then the first test seems to have no influence on the second. Boredom, and consequent carelessness, is a possible explanation. The first test leaves the person bored, and if the second test follows immediately, he will be careless. However, if the second test is sufficiently delayed, the person's boredom will dissipate and he will work carefully.

CONCLUSION

Split-half recognition tests appear to be mutually noninterfering if they are separated by 24 h or more. However, if they are separated by less than 1 min, then the first test may interfere with the second.

REFERENCES

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University, 1944.

UNDERWOOD, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 112-129.

(Received for publication November 22, 1977.)