JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR, 8, 463–470 (1969)

# Effects of Recall Tests on Long-Term Retention of Paired Associates[1]

GORDON A. ALLEN, WILLIAM A. MAHLER, AND W. K. ESTES[2]

*Stanford University, Stanford, California 94305*

Forty *S*s were given 5 or 10 paired-presentation training trials on paired-associate items followed immediately by 0, 1, or 5 unreinforced recall tests. The effects of the immediate test trials were to increase substantially long-term retention, as measured either by error frequency or response latency, to increase stereotypy of both correct and incorrect responses, and to reduce the frequency of intrusion relative to confusion errors. These findings were interpreted in terms of a distinction between storage and retrieval processes, the formation of associations being assumed to occur upon training trials and the availability, or retrievability, of the response members of items to increase as a function of occurrences on test trials.

Although the acquisition of paired associates is certainly a function primarily of informational reinforcing events supplied by the experimenter on each training trial, evidence is accumulating that learning in some sense may occur on recall tests, that is trials on which the subject is presented with the stimulus member of a pair and attempts to give the response from memory but receives no informative feedback. For example, Eimas and Zeaman (1963) found that latencies of correct responses decreased over successive recall tests. Izawa (1966) obtained evidence indicating that the short-term retention loss which normally follows a single training trial on a paired-associate item can be retarded or even prevented by a spaced sequence of recall tests. And a number of investigators (e.g., Goss, Morgan, & Golin, 1959; Richardson & Gropper, 1964; Butler & Peterson, 1965) have found upward trends in correct response probability over sequences of recall tests given after a series of training trials.

All of these results, however, have involved only relatively short-term effects and thus might be interpreted in terms of effects of test trials on the availability of associations in short-term memory. In this study we addressed ourselves to the question whether unreinforced recall tests produce learning as measured by long-term retention. In other words, we asked whether the effect of a recall test is simply to refresh an association, so to speak, in short-term memory, or whether it effects a transition from short-term to long-term memory storage. Our general approach was to obtain controlled comparisons of retention at a 24-hour interval after a training session in which paired-associate items received varying numbers of unreinforced recall tests immediately after paired-presentation trials. Also, the number of paired-presentation trials was varied orthogonally to the number of recall tests in order to determine the extent to which informative trials and recall tests trade off in their effects on long-term retention.

## METHOD

### *Subjects*

The *S*s were 40 college-age men and women; most were associated with Stanford University, and most had participated in other psychological experiments. Each *S* was tested individually and was paid $3.50 for the two sessions involved in the experiment, each session lasting approximately 30 min.

### *Apparatus and Materials*

The apparatus has been fully described by Izawa (1966). Briefly, *S* sat by himself in a sound-deadened,

[1] This research was supported in part by Grant GB-3878 from the National Science Foundation.

[2] Present address, to which reprint requests should be directed: The Rockefeller University, New York, New York 10021.

air-conditioned room which contained only the equipment necessary to display the stimulus and response (on training trials) or the stimulus (on test trials), and equipment on which *S* made his response on test trials. All control and recording apparatus was located in an adjoining room. The *S* sat in a chair 5.5 ft. away from the display box, on the front of which were an upper panel and a lower panel, identical except for position. The size of each panel was 2 × 12 in. Eight letters or digits could be displayed in each panel. The three letter positions of the left side of the upper panel were used for the stimulus presentations on both training and test trials. The response digits were shown on the two letter positions of the right side of the upper panel on training trials. The lower panel was not used in this experiment.

On the table in front of *S*'s chair was a horizontal panel containing two columns of response keys, each $\frac{7}{8} \times 1$ in. in size. Each column contained 10 response keys numbered 0–9 in ascending order, the 0 being closest to *S*. After the stimulus had appeared on the display panel on a test trial, *S* indicated his response by depressing first one key in the left-hand column and then one in the right-hand column, thus generating a two-digit number. As soon as *S* depressed a key, it was illuminated until the response had been recorded automatically in the adjoining room.

The stimulus members of the paired associates were 27 three-letter English nouns, taken from the highest frequency category in the Thorndike–Lorge tables of word frequency. The single set of 27 two-digit response numbers was selected randomly, subject to the restrictions used by Izawa (1966).

Subgroups of four *S*s received the same pairings of stimuli and responses. For each subgroup the stimuli were assigned randomly to the conditions, as were the responses; therefore, the pairings of stimuli and responses were also random. With nine conditions, there were three stimulus-response pairs per condition for each *S*.

*Design*

Training and testing involved two types of trials, which will henceforth be denoted R and T trials. An R trial on any item was a paired presentation of its stimulus and response members. A T trial was a recall test, on which the stimulus member was presented alone, *S* attempted to give the correct response, and no informative feedback was provided.

Three levels of training and three of initial testing were combined factorially to determine the nine experimental conditions for Day 1. Three items were assigned to each condition for each *S*. During the training phase of Day 1, items in Condition 10 were presented on 10 R trials; items in Condition 5F and 5L were presented on 5 R trials, all in the first half, or all in the second half of the training phase, respectively. Immediately after the training phase, one-third of the items of each training condition received five successive T trials, one-third a single T trial, and one-third no T trials. After a 24-hour interval, all items were tested on four successive T trials.

*Procedure*

At the beginning of the session on Day 1, *E* briefly explained what paired-associate learning is, what types of trials would be used, and how to respond on T trials. The *S* was told to repeat aloud the stimulus-response pair appearing on the screen as often as possible on R trials. It was emphasized that *S* should respond as quickly as possible on T trials since latencies would be recorded. The *S* was also told that there would be a short break of 5–10 sec. between the last R trial and the first T trial and that he should guess if he did not know the correct answer on a T trial since the next T trial would not begin until a response had been made.

The session began with 10 cycles of 18 R trials per cycle; nine items appeared in all 10 cycles, nine appeared in only the first 5 cycles, and nine appeared in only the last 5 cycles. The order within each cycle was randomized. Each stimulus-response pair appeared on the screen for approximately 2 sec., followed by an inter-trial interval of approximately 1 sec.

After the short break, the T trials began. The stimulus members of six of the items in each of the three training conditions were presented in random order; *S* was given as much time as he desired to respond, and his response and its latency, to the nearest .01 sec., were recorded. No information concerning correctness of responses was ever given once the T trials had begun. Three items from each training condition appeared on four additional cycles of randomly ordered T trials.

At the end of the session on Day 1, *S* was told to return 24 hr. later for "more of the same." He was not told whether the second day's session would use the same materials or not.

At the beginning of the session on Day 2, *S* was told that he would be tested on what he had learned on the previous day. The method of responding was briefly reviewed. The session comprised 4 cycles of T trials, each cycle being a random sequence of the 27 stimuli used in the experiment. Responses and latencies were recorded in the same manner as on Day 1.

## RESULTS

### *Error Data*

Proportions of errors for the group of 40 *S*s, by trials within each day and condition, are summarized in Table 1. In view of the randomization procedures, we can assume

that the items receiving no tests within each training condition on Day 1 would, if tested, have yielded error proportions approximately equal to those occurring on the first T trial of Day 1 for 1T and 5T items. Consequently it is apparent that there was a very large retention loss from Day 1 to Day 2 for items which were not tested on Day 1 and that this loss was largely independent of the training condition. Further, this overnight retention loss was substantially reduced by the effect of a single T trial after training on Day 1 and was almost completely eliminated by a sequence of five T trials after training on Day 1.

trials of Day 2 cannot be as satisfactorily evaluated since the analysis must be based on error scores with a range of only 0–3 (that is, the number of errors per *S* on any one trial over the three items assigned to the given combination of training and previous testing conditions). Nonetheless this analysis was done and the variation over Day 2 T trials proved significant at the .01 level. It does not seem, however, that this trend can represent learning in the same sense as the effects of original training trials or the effects of Day 1 tests upon long-term retention. It will be noted that the decrease in error proportions

TABLE 1

ERROR PROPORTIONS BY TRIALS AND CONDITIONS

| Number of training trials | Number of Day 1 tests | Day 1 test trial | | | | | Day 2 test trial | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 |
| 10 | 5 | .14 | .10 | .09 | .11 | .10 | .12 | .13 | .17 | .16 |
| | 1 | .07 | | | | | .19 | .18 | .17 | .15 |
| | 0 | | | | | | .35 | .28 | .32 | .28 |
| 5L | 5 | .10 | .12 | .12 | .11 | .09 | .18 | .17 | .12 | .14 |
| | 1 | .17 | | | | | .34 | .30 | .28 | .28 |
| | 0 | | | | | | .42 | .36 | .35 | .34 |
| 5F | 5 | .46 | .40 | .37 | .36 | .35 | .38 | .38 | .34 | .34 |
| | 1 | .39 | | | | | .45 | .45 | .42 | .43 |
| | 0 | | | | | | .63 | .61 | .58 | .55 |

Since each cell in Table 1 represents 120 observations, there is little doubt that the principal trends discernible in the table are reliable. To provide additional evidence on this point, we conducted an analysis of variance of the Day 2 test data, taking as the score for each subject, on each combination of training and previous testing conditions, the total number of errors made over the four T trials of Day 2 on the three items assigned to that combination of conditions. Effects of training conditions and number of Day 1 T trials were significant well beyond the .01 level but the interaction of these two variables was not significant. The significance of a slight downward trend in error proportions over the T

over T trials on Day 1 for the 5T conditions is followed by an overnight regression and then a decrease which in no case goes appreciably below, and in most cases does not reach the terminal level of Day 1.

With regard to the long-term effects shown in Table 1, it is of special importance to note that R trials and Day 1 T trials, though both increase long-term retention, are by no means interchangeable. For example, a combination of five R trials and five Day 1 T trials yields a much lower error probability on Day 2 than ten R trials with no Day 1 T trials. This last statement holds, of course, only for comparison of the ten with the 5L training condition. On the basis of the present experiment

alone one cannot say whether the substantial superiority of the 5L over the 5F training condition is due to retention losses which occurred for items of the 5F condition during the last five training cycles, when they were not presented and other items were undergoing learning, or whether it is the result of a learning-to-learn effect such that the early training trials were much less effective than later training trials in producing learning. Other data from our laboratory (for example, see Mahler, 1968) indicate that almost certainly the former interpretation is substantially correct.

The manner in which the initial tests influence long-term retention can be elucidated somewhat by a breakdown of Day 2 errors into two main types. Only 27 of the possible two-digit numbers occurred as response members. Denoting these as the "response set," we can break down Day 2 errors into those in which the response that was given belonged to the response set (confusion errors) and those in which the response given did not belong to the response set (intrusion errors). The relative proportions of the two types of errors to be expected on the basis of chance for items on which no associative learning had occurred cannot be specified precisely, owing to the restrictions employed in drawing the response sets. On the basis of an inspection of the protocols, it appears that the subjects recognized that responses of the form 0i, and ii were not used as response terms and almost never used them in guessing. Thus, if all errors represented random guesses from the remaining 80 possible two-digit numbers, approximately 34% of the errors should have been confusion errors. The overall percentage of Day 2 errors which fell in the confusion category was 45, which is substantially in excess of chance expectation on any basis.

The full breakdown by types of errors and Day 1 conditions is given in Table 2. The most conspicuous differential trend is the relative constancy of confusion errors as a function of number of Day 1 tests in contrast to the strong negative correlation between the number of Day 1 tests and proportion of intrusion errors. Looked at differently, 56% of the errors in the 5T condition fell in the confusion category but only 40% of the errors in the 0T condition. It does not appear that the simple factor of response availability, which might be expected to be directly related to the frequency with which the responses belonging to items of the various testing conditions occurred on Day 1, has anything substantial to do with the observed pattern of Day 2 errors. When the Day 2 confusion errors are broken down according as the responses given belonged to the 5T, 1T, or 0T subsets, it is found that almost precisely one-third fell in each category (gross frequencies of 202, 205, and 204, respectively).

TABLE 2

PROPORTIONS OF CONFUSION AND INTRUSION ERRORS ON DAY 2

| Number of training trials | Number of Day 1 tests | Confusion errors | Intrusion errors |
|---|---|---|---|
| | 5 | .10 | .04 |
| 10 | 1 | .09 | .08 |
| | 0 | .13 | .18 |
| | 5 | .07 | .08 |
| 5L | 1 | .13 | .17 |
| | 0 | .16 | .21 |
| | 5 | .19 | .17 |
| 5F | 1 | .18 | .26 |
| | 0 | .21 | .38 |

Perhaps the simplest interpretation of this pattern of results is that confusion errors in many instances represent, not absence of learning, but cases of learned items on which errors occur because of similarities between items with respect to either stimulus or response properties. Intrusion errors, on the other hand, may represent primarily instances when S was entirely unable to retrieve a learned response upon presentation of a stimulus and simply had to guess. If this interpretation is correct, then the effect of Day 1 T trials is primarily to increase retrievability of the

response members of items which were learned during the R series on Day 1.

A different, though not independent, way of examining retention loss over the 24-hour interval is to compute the relative frequencies with which *S*s switched from correct responses on Day 1 tests to errors on Day 2 tests for particular items in the 1T and 5T conditions. These data are presented in the upper half of Table 3 in terms of the conditional proportions of errors on the first trial of Day 2 for items on which the final test of Day 1 yielded a correct response. This index of retention loss, though only weakly related to training conditions, is very substantially affected by number of Day 1 T trials, the level for items having had five Day 1 tests being less than half of that for items receiving only one Day 1 test. The proportions of errors on the first trial of Day 2 given an error on the same item on the last test of Day 1, shown in the lower half of Table 3, are not entirely comparable since they do not reveal the extent to which *S*s may have changed from one erroneous response to another on a given item over the interval. A further breakdown of the conditional error frequencies in this respect again shows a large effect of Day 1 tests; for 50% of the 5T items on which errors occurred on the last test of Day 1, the same error recurred for the same item at least once on Day 2, whereas the corresponding percentage for 1T items was only 18.

TABLE 3

CONDITIONAL PROPORTION OF ERRORS ON TRIAL 1, DAY 2, GIVEN CORRECT OR ERROR ON LAST TRIAL, DAY 1

| | Number of training trials | Number of Day 1 test trials | |
|---|---|---|---|
| | | 1 | 5 |
| Error given correct | 10 | .16 | .06 |
| | 5L | .24 | .10 |
| | 5F | .20 | .09 |
| Error given error | 10 | .62[a] | .67[a] |
| | 5L | .91[a] | .82[a] |
| | 5F | .83 | .91 |

[a] $N < 40$.

Thus, there was substantially more stereotypy of response over the retention interval after five Day 1 tests than after a single Day 1 test and the effect was of about the same magnitude for errors as for correct responses. The more often an item is tested immediately after training, the more likely it is that whatever response is made to the item on the T trials will be repeated after a long retention interval. It might be remarked in this respect that examination of the individual protocols reveals a large number of instances in which *S*s settled upon a particular error for a particular item in the 5T condition and made this error repeatedly over the later trials of the test series and also a considerable frequency of the same phenomenon for all conditions over the sequence of tests on Day 2. In view of the fact that this type of stereotypy on Day 2 is substantially related to the number of previous tests on Day 1, we evidently must conclude that the results arise at least in part from some form of learning which occurs on the T trials and not simply from pre-existing associations between stimulus and response members of the items.

*Latency Data*

Response latencies, computed separately for correct responses and errors, are presented in Table 4 for each trial in relation to training and testing conditions. It does not seem feasible to do any overall statistical analysis of the correct and error latencies separately in view of the large variation in number of observations from cell to cell. However, an analysis of variance for the pooled mean response latencies on Day 2 as a function of training and Day 1 testing conditions shows all of the main effects and the interaction of these variables to be significant far beyond the .01 level. Since differences between means for the various training and testing conditions are of about the same order of magnitude for the correct and error latencies, it seems safe to

TABLE 4

MEAN LATENCIES BY TRIALS AND CONDITIONS

| Number of training trials | Number of Day 1 tests | Correct response latencies | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Day 1 test trial | | | | | Day 2 test trial | | | |
| | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 |
| 10 | 5 | 2.31 | 2.11 | 1.68 | 1.71 | 1.56 | 2.03 | 1.81 | 1.90 | 1.53 |
| | 1 | 2.04 | | | | | 2.29 | 1.82 | 1.76 | 1.71 |
| | 0 | | | | | | 2.50 | 2.12 | 1.81 | 1.50 |
| 5L | 5 | 2.07 | 1.82 | 1.71 | 1.61 | 1.60 | 2.13 | 1.60 | 1.61 | 1.48 |
| | 1 | 2.28 | | | | | 2.74 | 1.97 | 1.84 | 1.68 |
| | 0 | | | | | | 2.62 | 2.00 | 1.85 | 1.54 |
| 5F | 5 | 2.79 | 2.03 | 1.87 | 1.64 | 1.54 | 1.86 | 1.70 | 1.69 | 1.49 |
| | 1 | 2.85 | | | | | 2.56 | 2.10 | 1.91 | 1.81 |
| | 0 | | | | | | 2.77 | 2.22 | 2.15 | 2.08 |
| | | Error latencies | | | | | | | | |
| 10 | 5 | 4.86 | 3.10 | 3.81 | 2.59 | 2.37 | 3.80 | 4.09 | 3.96 | 3.11 |
| | 1 | 6.00 | | | | | 5.05 | 4.26 | 2.81 | 2.69 |
| | 0 | | | | | | 4.68 | 4.32 | 3.64 | 3.17 |
| 5L | 5 | 5.80 | 5.18 | 4.53 | 2.44 | 2.18 | 3.49 | 2.89 | 3.15 | 2.35 |
| | 1 | 4.18 | | | | | 4.43 | 4.13 | 3.87 | 3.63 |
| | 0 | | | | | | 5.00 | 4.43 | 3.80 | 3.50 |
| 5F | 5 | 4.89 | 4.44 | 3.32 | 3.23 | 2.77 | 3.24 | 3.08 | 3.41 | 2.96 |
| | 1 | 4.60 | | | | | 4.58 | 4.52 | 3.30 | 3.12 |
| | 0 | | | | | | 5.40 | 4.29 | 3.63 | 3.62 |

conclude that all of the principal trends discernible in Table 4 for the correct and error latencies separately are quite reliable.

An overall pattern which may prove to be of major theoretical significance is to be seen in the ordering of mean latencies at the beginning of Day 2 as a function of training and Day 1 testing conditions. This order, for both correct and error latencies, closely parallels the ordering for error proportions seen in Table 1. As was found for the error proportions, there is relatively little variation in Day 2 latencies as a function of training conditions, but major variation in relation to number of Day 1 tests. The similarity in pattern for the correct and error latencies is quite striking; it may be noted that, throughout the table, mean error latency for any combination of conditions is approximately twice the corresponding correct response latency. The principal differences in trends between latencies and error frequencies are that the former exhibit a rather greater decline within days and a greater increase from the end of Day 1 to the beginning of Day 2 than might have been expected from the frequency data, and a somewhat greater convergence over the T trials of Day 2 for the mean latencies representing different Day 1 testing conditions.

## DISCUSSION

There seems to be no doubt concerning the answer to the principal question at issue in this study. Relatively long-term retention of paired associates is substantially influenced by recall tests given immediately after training. Within any sequence of trials, the immediately observable effect of a T trial is a reduction in response latency, and this reduction seems to be of about the same magnitude regardless of

whether the response is correct or incorrect. The decline in latency which occurs over a series of closely spaced T trials is followed by regression to a point intermediate between the initial and terminal levels of the first series after a 24-hour rest interval, then during a second test series by another decline to approximately the same terminal level. Analysis of different types of errors suggests that learning that occurs on T trials operates to prevent failures of retrieval, and thus to lower the incidence of intrusion errors, on subsequent tests, but has little effect on the incidence of confusions. It might be remarked that all of the principal trends in the present data agree with those observed in an unpublished pilot study conducted by the writers, which was of similar scope but utilized somewhat more difficult material, and with the results of a recent study by Mahler (1968) that utilized a short period of interpolated learning rather than an overnight interval between the initial and terminal test series.

Perhaps the most parsimonious interpretation of the learning which occurs on unreinforced recall tests would be that it is basically the same as that occurring on paired-presentation training trials, the only difference being that on tests the occurrence of the response member of the paired-associate item is under *S*'s control. Evidently this simple interpretation is not adequate, however, for R and T trials prove not to be interchangeable in their effects on retention as measured either by error probabilities or latencies. For example, it is clear in both Table 1 and Table 4 that five R trials plus five immediate T trials produce long-term retention much superior to that observed after ten R trials with no immediate tests. Even more strikingly, the addition of a single test after ten R trials reduces error frequency after a 24-hour interval by 50% as compared to ten R trials without the immediate test.

The pattern of results appears to fit in rather well with the distinction between storage and retrieval processes in paired-associate learning which was suggested on the basis of a quite different type of evidence in an earlier study by Estes & Da Polito (1967). The principal basis for the distinction in that study was the finding that the amount of information stored in memory, as measured by a recognition test, was approximately equal after intentional versus incidental training procedures, whereas recall performance was drastically impaired after incidental training. In the present study it appears that the availability, or retrievability, of the response member of a paired-associate item increases as a direct function of its frequency of occurrence on T trials. This conception would account, not only for the effects of early T trials on long-term retention, but for the parallel changes in correct and error latencies, and for the similar effects of T trials on stereotypy of correct responses and errors. Whether learning in the sense of an increase in the long-term retrievability of the response member of a paired-associate item occurs also on paired-presentation trials is not clear. Pending more direct evidence, the simplest interpretation would seem to be that it does not occur on paired presentations per se but may occur during rehearsal immediately after paired-presentation trials, though in less effective fashion than on recall tests when the stimulus member of the item is present.

The trends in our latency data differ in one major respect from those reported by Eimas and Zeaman (1963). Whereas in both studies correct response latency decreased substantially over successive T trials, in our data error latency decreased similarly but in Eimas and Zeaman's data error latency was virtually constant. The only plausible explanation that has occurred to us has to do with the categorization of errors as repetitive or nonrepetitive. In Eimas and Zeaman's study, there were only a few cases in which a subject made the same incorrect response to a given item on successive T trials, and in this portion of the data mean latency decreased slightly from the first to the second T trial. For the remainder

of their error data, involving nonrepetitive errors, latency was constant over tests.

A similar breakdown of our Day 2 latency data (pooled over conditions to obtain adequate $N$'s) is presented in Table 5. For the cases in which the same error occurred on all four tests, latency declined fully as steeply as for correct responses. But for the items represented in the third row of Table 5, on which Day 2 responses were all incorrect but not all the same error, the function is appreciably shallower. Further, if these means are converted to reciprocals, the change from the first to the second test for the "Different Error" category is very slight. Thus there may actually be no appreciable disparity between the corresponding trends in these two studies.

TABLE 5

MEAN DAY 2 LATENCIES FOR ITEMS WITH ALL CORRECT RESPONSES, ALL SAME ERRORS, OR DIFFERENT ERRORS

| | | Trial | | | |
|---|---|---|---|---|---|
| Type | $N$ | 1 | 2 | 3 | 4 |
| All C | 620 | 2.20 | 1.74 | 1.64 | 1.55 |
| Same E | 54 | 4.07 | 3.05 | 2.70 | 2.26 |
| Diff. E | 406 | 4.36 | 3.79 | 3.37 | 2.93 |

The substantially steeper decline in latencies for repetitive errors seems to fit well with the assumption that, to a major extent, paired-associate latency reflects the state of retrievability of the stimulus-response association. Other things equal, retrievability of a given response varies directly with frequency and recency of evocation of that response (whether correct or incorrect) by the given stimulus, and is to be distinguished from the notion of response availability (Horowitz, Norman, & Day, 1966; Underwood, Runquist, & Schulz, 1959) which is not stimulus-specific.

## REFERENCES


BUTLER, D. C., & PETERSON, D. E. Learning during "extinction" with paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 103–106.

EIMAS, P. D., & ZEAMAN, D. Response speed changes in an Estes' paired-associate "miniature" experiment. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 384–388.

ESTES, W. K., & DA POLITO, F. Independent variation of memory storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 18–26.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *Journal of Experimental Psychology*, 1959, **57**, 96–104.

HOROWITZ, L. M., NORMAN, S. A., & DAY, R. S. Availability and associative symmetry. *Psychological Review*, 1966, **73**, 1–15.

IZAWA, C. Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, 1966, **18**, 879–919.

MAHLER, W. A. Effects of study and test trials on retention of paired associates. Unpublished Masters Thesis, Stanford University, 1968.

RICHARDSON, J., & GROPPER, M. S. Learning during recall trials. *Psychological Reports*, 1964, **15**, 551–560.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.


(Received December 11, 1968)