1972

Recognition and recall as measures of retention on a paired associate task

David F. Prim

RECOGNITION AND RECALL AS MEASURES OF RETENTION ON A PAIRED ASSOCIATE TASK

David F. Prim

A thesis submitted in partial fulfillment
of the requirements for the degree of Master of Arts
in psychology in the Graduate School of the
University of Richmond

July, 1972

RECOGNITION AND RECALL AS MEASURES OF
RETENTION ON A PAIRED ASSOCIATE TASK

David F. Prim

Approved: Dr. L. James Tromater

Dr. Kenneth A. Blick

Dr. Frederick J. Kozub

ACKNOWLEDGMENTS

I gratefully acknowledge the professional talent and assistance that was given by Dr. James Tromater and the members of my thesis committee. Without their guidance, this project would have never realized fruition.

I also acknowledge a source of even greater assistance, my wife, Joan. She endured countless hours of loneliness, provided much needed moral support, and gave freely of her typing abilities.

RECOGNITION AND RECALL AS MEASURES OF RETENTION ON A PAIRED ASSOCIATE TASK

David Prim

University of Richmond

Abstract

Widely disparate findings concerning recognition and recall as indicants of retention have been reported by several independent researchers. To clarify the problem a list of 8 items, composed of letter-number pairs, was presented 5 times by the study-test method to 160 college undergraduates. The list was learned by either recognition or recall and then tested by either a recognition or recall test after 24 hour and 72 hour intervals. *S*s were placed in 1 of 5 categories dependent upon the trial the *S* achieved 100% criterion. A 4 factor ANOV showed recognition scores to be significantly higher at the .05 level than recall scores.

The measurement of retention has intrigued, fascinated, and confounded investigators since the classical study of Ebbinghaus (1913). His attempts to experimentally quantify retention and investigate higher mental processes generated areas of research that continue today. C. W. Luh (1922) published a now famous monograph which established the body of information that was the authoritative reference on retention measures until 1957 when Postman and Rau compiled and published a report comparing measures of retention. Postman and Rau began their investigation with the statement, "The one fact for which there is substantial experimental evidence is that tests of recognition yield higher scores than do tests of recall[p.218]." This statement was relatively safe from challenge until 1964 when Bahrick asserted that, " conclusions regarding the superiority of recognition to recall performance, and regarding the slope of retention curves are overgeneralizations, and therefore misleading, because the findings on which they are based do not represent intrinsic differences between indicants of recognition and recall [p. 188]." These diametrically opposed statements provide a framework for investigation since other experimenters have chipped away at the differences in recognition and recall measures with good success. This study was conducted to investigate the validity of Bahrick's assertions in light of experimental evidence accumulated since 1964.

Bahrick's statement concerning conclusions based on differences between recall and recognition measures is based on the premise that artifacts in design, overlearning, and easy recognition tests unduly inflate the recognition scores. According to Bahrick, the correct design for comparing retention for recall and recognition is to train individual subjects (Ss) until all of their recall responses are correct, and another group of individual Ss until all of their recognition responses are correct. Previously, investigators had given all Ss a constant number of training trials and later compared performance on recognition and recall tasks.

When the objective of the experimental effort is to examine the test rather than the stimulus materials it is necessary to bring each group to comparable criterion on the same task before administering the test. The degree of original learning with respect to number of reinforced trials must be equated before any valid statement can be made concerning differences between the test measures.

Underwood (1964) in an attempt to popularize his single and multiple entry projection techniques argued that performance to a criterion is not a valid measure of degree of learning. Concerning criterion performance on lists of different difficulty Underwood states, "it has often been assumed that degree of learning was equivalent and that, therefore, differences in retention reflect the effect of some other variable. This assumption cannot be justified. Logically, we must expect that when acquisition curves approach a common criterion at different rates, and the learning is stopped at this criterion, the projection of the curves for one additional trial cannot result in equivalent performances [p. 122]."

In any eventuality it is clear that the need to equate or control degree of original learning is paramount if a learning/performance distinction is to be made. If the original learning is not equated or otherwise controlled, no definitive statements concerning the differential effects of performance on recognition or recall tests can be made.

A classical experiment by Krueger (1929) points out the effects of even a small degree of overlearning on performance. Using a list of 12 nouns as learning material and retention intervals from 1-28 days, Krueger found recall and savings scores increased rapidly at first as degree of overlearning was varied from 0-100%. Krueger's results may be severely vitiated by proactive interference since his Ss served in several conditions of the experiment and were well practiced. Postman (1962) investigated relearning and recall as a function of degree of overlearning. Using serial lists of high and low frequency words, Postman found that the amount recalled showed a positively accelerated increase with degree of overlearning. The facilitation in the recall measure was largely due to improved retention of difficult items in the lists. Postman used naive Ss who learned and recalled a single list. Where there is a large amount of proactive interference it appears that practically all items will have to be overlearned if they are to be recalled.

Postman's conclusions regarding the amount of overlearning required for recall of easy and difficult items has been challenged by Greenfield (1969). Greenfield, using 16 syllable-noun pairs conducted two experiments using recognition and recall as indicants of retention. Greenfield concluded overlearning increases associative strength for both hard and easy pairs and that when the pairs are overlearned in the same condition they increase equally in associative strength.

Bahrick (1964) discussed the impact of overlearning on retention measures and concluded that "indicants of retention are not sensitive to early retention loss if the material has been overlearned with respect to the threshold of that indicant [p. 190]." To examine the effects of overlearning on recognition it is best to examine those instances where training stopped near the recognition threshold. Strong (1913) did this and reports a negatively accelerated curve for recognition scores. In general, overlearning tends to make material less vulnerable to interference and as such differentially affects measures of recall and recognition since recognition does not require production of the response, only differentiation.

Various models of memory and recall postulate a dual process theory to account for differences between recognition and recall. Estes and DaPolito (1967) investigated the effects of incidential versus intentional learning instructions as measured by recognition and recall tests. They found little decline in performance on recognition tests under either set of instructions but recall measures showed a large performance decrement under the incidential learning condition. The authors invoked a concept of rehearsal under the intentional instructions condition which would modify recall scores by placing some items over threshold. Davis and Okada (1971) investigated recognition and recall performance for individually cued words which Ss were to either remember or forget. They found that Ss retained words they were instructed to remember. The reason cited for the differential recall was not rehearsal as one might expect. A concept of blocked or inferior retrieveability was invoked to explain the poorer retention of "forget" items. Bjork (1970) tends to favor rehearsal as an answer for lack of durability of "forget" items. He contends that forget instructions effectively reduces rehearsal which in turn results in the formation of fewer retrieval cues.

Loftus (1971) found differences in storage procedure between recognition and recall. Loftus varied the Ss knowledge at the time of study of how he would be tested. It was found that knowledge of test measure increased recall performance but did not similarly increase recognition performance.

Butterfield, Belmont, and Peltzman (1971) present further evidence of facilitation of recall by knowledge of test method. The authors manipulated memory demand by varying the response requirement and examined the extent to which Ss used rehearsal. They observed that when Ss have prior knowledge about the recall requirement they recall more than when cued after acquisition. From the preceding studies it appears that the prior knowledge of method of retention test facilitates recall and has little effect of recognition.

Kintsch (1968) provided data indicating that organization of stimulus material facilitated recall but had little effect on recognition. Kintsch demonstrated that organization in terms of conceptual categories is not an important variable in recognition but has a pronounced effect on recall. Kintsch interpreted the data in favor of a dual process retreival model similar to that of Estes and DaPolito. Bruce and Fagan (1970) extended Kintsch's study and supported his findings. They further demonstrated that failure to find significance of organization in the recognition mode was not due to an easier recognition test. Numerous other investigators (Lewis 1971; Luek, McLaughlin, & Cicala 1971; Wood 1969) have found differences between structured and nonstructured lists and the difference appears to be reliable.

Postman, Jenkins, and Postman (1948) varied the sequence of test presentation to determine if there are significant effects. One group received training on nonsense syllables followed by a recognition then recall test. The second group received the same training except they received a recall test followed by a recognition measure. The authors reported recognition to be poorer after recall than before and that recall is better after a recognition test than before. Apparently the recognition test in effect served as additional learning for those in the recall group. Possibly some items that were just beneath recall threshold were strengthened enough by their appearance on the recognition test to boost them over the threshold.

Darley and Murdock (1971) in an attempt to clarify the nature of a negative recency effect found by Craik provided data concerning the effects of prior recall testing on final recall and recognition. Darley and Murdock presented each S ten lists of words followed by either a free recall test or no test at all. The Ss then received a final recall or recognition test on the words from all ten lists. They found that initial testing facilitates retrieval for recall for all serial positions but had no overall effect on recognition performance. The authors concluded that prior testing increased item accessibility but not availability. From the preceding studies it is concluded that recall performance is facilitated by prior testing, be it recall or recognition.

Deese and Hulse (1967) illustrate one difficulty in constructing recognition tests. The degree of difference between the incorrect and correct responses determines the difficulty of the test. If the alternate incorrect items are dissimilar to the correct item the test is judged to be very easy and scores will be high. Postman, Jenkins, and Postman (1948) constructed recognition items consisting of the correct nonsense syllable, a syllable with a one letter change from the correct one, and two additional distractor syllables which differed from each other by only one syllable. They found their Ss chose the incorrect syllable with two letters in common with the correct one a significant percentage more than the other two items.

Postman (1951) found that results of recognition tests varied inversely with the number of letters common to correct and incorrect alternatives on the recognition test. The more elements common to both, the greater the degree of difficulty of that item. When the incorrect alternatives are very similar to the item originally learned the S has to learn the whole item, just as in a recall mode, to discriminate between the similar alternatives.

The effect of degree of differentiation of alternatives has not received a great deal of investigation; however, the data suggest that the threshold required for recognition may be increased or decreased by manipulating the degree of similarity of item alternatives.

Just as similarity of response elements affects performance, the number of possible responses in a set acts to influence recognition performance also. On a test where four possible responses are given the S confines his attention to those four only and selects the one that he recognizes. For the comparable task on a recall test the S must choose among all the possible responses of which he has knowledge.

Davis, Sutherland, and Judd (1961) analyzed information content in recognition and recall where the number of alternatives was fixed. Davis et al. devised lists of 15 two digit numbers and 15 two letter syllables and tested by recall or recognition. Each S served in four conditions: recognition out of a list of 30, recognition out of a list of 60, recognition out of 90, and recall from 90. Under these conditions it was found that the amount of information transmitted was not significantly different.

Grasha, Reichmann, Newman, and Fruth (1971) studied the situation in which the response sets for recognition and recall were equated and available. Using a one trial procedure with seven or nine consonants as material the authors found no significant difference between recognition and recall.

McNulty (1965) hypothesized that differences between the measures may be due in part to the use of the whole item as the unit of measurement. McNulty asserts that some Ss learn less than the whole item and on the basis of this partial learning are able to recognize but not recall the item. Using approximations to English as stimulus materials McNulty found the differences between recall and recognition disappeared when partial learning was controlled. In this experiment the recognition test alternatives varied from the original item by only one letter out of eight.

The extensive analysis by Postman and Rau appears to have been effectively criticized by several experimenters. Bahrick's assertions have received too much support to ignore, but not enough direct examination to support it in its entirity. No single experiment has been conducted which incorporated the design suggested by Bahrick with proper controls for overlearning, instructions, knowledge of test method, number of alternatives, and organization of material. The null hypothesis of no difference between recognition and recall is tested by comparing performance on each test measure when the independent variables are controlled.

METHOD

Design

A 5x2x2x3 factorial design with repeated measures on the last factor enabled the testing of 5 levels of original learning (factor A), under two learning methods, recognition and recall (factor B), measured by two indicants of retention, recognition and recall (factor C), over a period of 24 and 72 hours (factor D). The third measure included in Factor D was the score each individual S achieved at the end of the last trial. Two prior pilot studies demonstrated that degree of overlearning was very difficult to control under the best of circumstances, therefore, overlearning was incorporated into the design as a category factor. A frequency plot of trials to criterion (TTC) showed that Ss divided themselves between trials 2 and 5 with an additional category, 5+, added for those Ss who had not achieved criterion at the end of the fifth trial. Category 1 included Ss who achieved criterion on trial 2, category 2 encompassed those Ss who reached criterion on trial 3 and so forth through trial 5+. The number of items correct at the end of the last trial, the number retained after 24 hours and the number retained after 72 hours were used as the dependent variable.

Subjects

Ss were 160 naive male and female undergraduate students attending the summer session at the University of Richmond. Only that data from Ss who completed all 3 test sessions were used for analysis. Data from Ss who indicated they had participated in a learning experiment within the preceding calendar year were excluded.

Apparatus

A 35mm Kodak Carousel projector, was used to project letter-number pairs at 5 second intervals on beaded projection screens. The slides consisted of white numbers and letters on a black velvet background. Instructions were recorded on a Lloyds cassette portable tape recorder. A Chesterfield Dolmy stopwatch was used to measure time lapse for retention tests.

List and Test Construction

Eight two-digit numbers were paired with letters of the alaphabet to provide list content. The numbers were selected to insure there were no forward sequences such as 23, 45, 67; no double digits; and each integer appeared only once in the first and second positions. The resulting list spanned from 28-97. Meaningfulness of selected numbers, as measured by associative value, Battig and Spera (1962) ranged from .88 for 59 to 1.69 for 28 with a mean of 1.31 for all eight numbers. Letters from the alphabet were chosen to limit possible acoustical interference even though the numbers are not to be pronounced out loud. Letters that rhymed or contained "ee" sound were excluded from consideration. The meaningfulness of the selected letters as measured by associative value, Anderson (1965), averaged 11.14 with a range from 8.80 for the letter K to 12.2 for the letters H and N. The letters and numbers were randomly paired, resulting in the following list:

H 61, N 43, L 86, K 97, W 59, Q 35, R 72, and F 28. Five separate random sequences of the list were constructed to vary the serial position of the items. The words START and STOP preceded and concluded each trial. The recognition test consisted of the presentation of the stimulus letter with four numerical alternatives. The alternatives consist of the correct number, a number from within the list, and two double digit distractor numbers chosen at random. The position and sequence of alternatives were varied randomly from trial to trial. The stimulus letters were randomly varied with the provision that they not occupy the same serial position as in the sequence displayed on the screen. In order to equalize the tasks the recall tests consisted of the same random sequence of letters as the recognition tests, but without the alternatives. The final recognition and recall tests displayed the same sequence of letters but that sequence was different from any of the preceding trials. Recall and recognition test booklets consisting of a page of instructions and five trial sheets were used. Following each trial answer sheet there was a page advising the S to not turn that page until further instructions were received.

Procedure

Test booklets were distributed face down to each S until all Ss had received a booklet. Tape recorded instructions were played advising that course grade, class standing, etc. would not be influenced by the outcome of the experiment and that individual results will be held confidential. Each S was then instructed to follow along by reading the instructions on the face of the test booklet. The instructions for the recognition and recall booklets were identical. Each S was informed of task requirement, the presentation rate, and the number of items. The work "START" was projected on the screen 5 seconds before each trial. "STOP" concluded each list and served as a cue to begin the test phase. The type of test to be administered after each trial was not divulged. Each item was displayed for five seconds. At the conclusion of each trial the Ss were instructed to turn the page and records their answers. Both recall and recognition tests were allocated 30 seconds for completion. After five trials had been administered the booklets were collected and the original learning session was terminated. No mention was made of the intent to return later for retesting. Twenty-four hours and again 72 hours after original testing a second and third recall or recognition test was given.

RESULTS

An unweighted means technique, employing the harmonic mean, was used in analysis as the number of Ss for factor A were unequally divided among the five levels. Forty Ss were used in each treatment condition, recognition-recognition, recognition-recall, recall-recongnition, and recall-recall, producing a total of 480 observations since each S was observed under three retention intervals.

An analysis of variance (ANOV) of the four factors, category x learning method x test method x retention interval, produced significant F ratios for several factors and interactions. Table 1 presents a summary of the ANOV.

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

The overall effects of the category factor (A) were significant, $F(4, 140) = 5.93$ $p < .01$. The significant F of the overlearning factor is not surprising nor unanticipated. A Newman-Keuls test of ordered means was performed on the means of factor A and a summary of the results is depicted in Table 2.

- - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - -

The means align themselves as a direct function of the number of reinforced trials after reaching criterion. The mean of category 5, reflecting scores from those Ss who required more than five trials to reach criterion, was significantly lower than all the other category means. The mean of category 4 was significantly lower than the means of categories 1 and 2. There were no significant differences between categories 1, 2, and 3. In each instance significance was judged on the basis of a comparison of the difference with a critical value computed from the Studentized Range Statistic. The interaction between category (A) with retention interval (D) was statistically significant, $F(8, 280) = 3.93$ $p < .01$. An

TABLE 1

Analysis of Variance: Category X Learning Method X Test Method X Retention Interval

Source	df	MS	F
Between subjects	159		
Category (A)	4	130.82	5.93**
Learn method (B)	1	21.09	
Test method (C)	1	128.76	5.84*
A X B	4	3.52	
A X C	4	5.10	
B X C	1	25.60	
A X B X C	4	1.49	
Subj w. groups (error between)	140	22.04	
Within subjects	320		
Interval (D)	2	76.08	43.47**
A X D	8	6.89	3.93**
B X D	2	10.01	5.72**
C X D	2	33.08	18.90**
A X B X D	8	.83	
A X C X D	8	1.49	
B X C X D	2	6.68	3.81*
A X B X C X D	8	.54	
D X subj w. groups (error within)	280	1.75	

*p < .05

**p < .01

TABLE 2

Newman-Keuls Test of Differences Between

All Pairs of Means of Category Factor (A)

Category		5	4	3	2	1
	Means	4.26	6.09	7.00	7.42	7.70
5	4.26		1.83*	2.74*	3.16*	3.44*
4	6.09			.91	1.33*	1.61*
3	7.00				.42	.70
2	7.42					.28
1	7.70					
			r = 2	r = 3	r = 4	r = 5
	$g_{.95}(r, \infty)$		2.77	3.31	3.63	3.86
$\sqrt{MSerr/\tilde{n}}$[a]	$g_{.95}(r, \infty)$		.91	1.09	1.19	1.27

[a]MSerr = 8.515 ñ = 78.125

*p <.05

analysis of simple effects of all interactions was computed and is presented in summary form in Table 3. Retention interval was significant for all levels

- - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - -

of factor A. The profile of the AXD interaction is presented in Figure 1.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

An examination of the AXD profile indicates the source of interaction is between the 24 hour and 72 hour intervals for all levels of factor A. At level a_2 the 24 hour measures yielded the higher mean whereas at level a_3 the 72 hour mean was higher. At level a_4 the 24 hour mean was again higher.

The F ratio for the main effects of the test method (C) was also significant, $F(1,140) = 5.84$ $p < .05$ as was the main effects of retention interval (D), $F(2,280) = 43.47$ $p < .01$. Factor B, the test method, failed to reach significance and provided an $F < 1$. Therefore, comparisons can be made between test results obtained from these two methods of learning.

The interpretation of the significance of the main effects of factor C is clouded by the significant interactions of test method with interval(CXD), $F(2,280) = 18.90$ $p < .01$ and the three factor interaction of learning method with test

TABLE 3

Analysis of Variance: Simple Effects of Significant Interactions

Source of Variation	df	MS	F
Interval at category (A x D)			
d at a_1	2	6.11	3.49*
d at a_2	2	22.59	12.91**
d at a_3	2	68.34	39.05**
d at a_4	2	254.58	145.47**
d at a_5	2	63.16	36.09**
D x subj w groups	280	1.75	
Learning method at intervals (B x D)			
b at d_1	1	9.76	
b at d_2	1	221.37	25.99**
b at d_3	1	180.03	21.14**
Within cell	420	8.51	
Test method at intervals (C x D)			
c at d_1	1	.073	
c at d_2	1	925.90	108.73**
c at d_3	1	1023.38	120.18**
Within cell	420	8.51	
Learn, test method at intervals (B x C x D)			
BC at d_1	1	4.94	
BC at d_2	1	704.67	82.75**
BC at d_3	1	665.57	78.16**
Within cell	420	8.51	

* $p < .05$

** $p < .01$

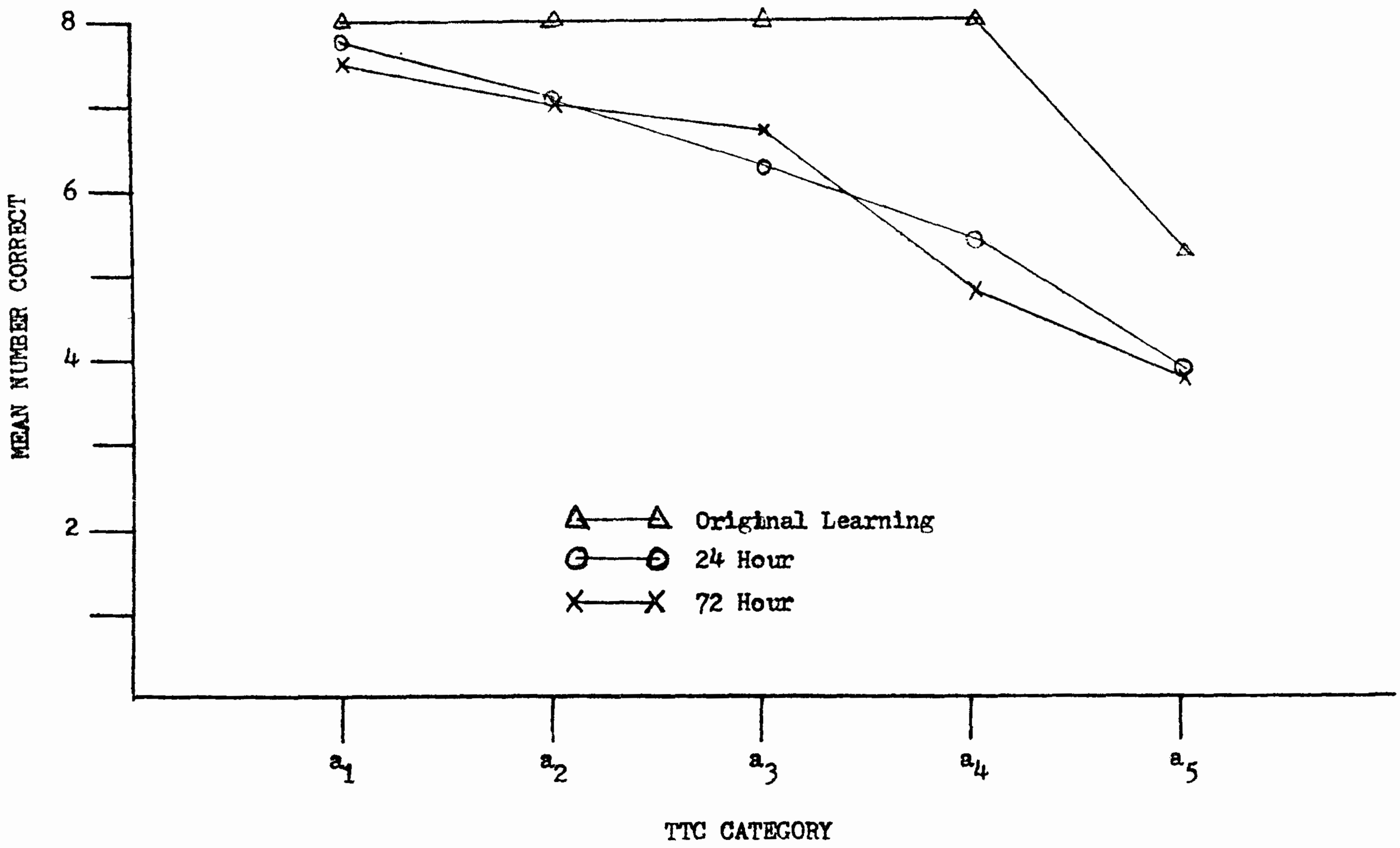

Fig. 1. Profile of interaction of category factor with retention interval (AXD).

method with retention interval (BXCXD). There was also a significant interaction between learning method and retention interval (BXD), $F(2,280) = 5.72$ $p < .01$. An examination of the profiles for CXD and BXD in Figure 2 show that the inter-

- - - - - - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - - - - - -

actions result from higher recall scores at interval d_1 in both cases. The profile lines are practically parallel for intervals d_2 and d_3 for both interactions. The simple effects analysis in Table 3 indicate no significant differences between b at d_1 and c at d_1. These two interactions indicate equality at time of original learning rather than a true mixing of treatment effects. The difference between learning methods at 24 hours and 72 hours is very real and significant producing ratios of $F(1, 420)$ of 25.99 and 21.14, $p < .01$. Those Ss who learned under the recognition method were performing significantly better than those Ss learning under the recall method. Differences in test measures at d_2 and d_3 also were highly significant, $F(1, 420) = 108.72$ $p < .01$ and $F(1, 420) = 120.18$ $p < .01$ respectively. The scores obtained by the recognition tests were significantly higher than those obtained by the recall tests.

The BXCXD interaction is shown in profile form for each level of d in Figure 3. The parallel lines at d_1 show no significant interaction; the analysis of simple effects

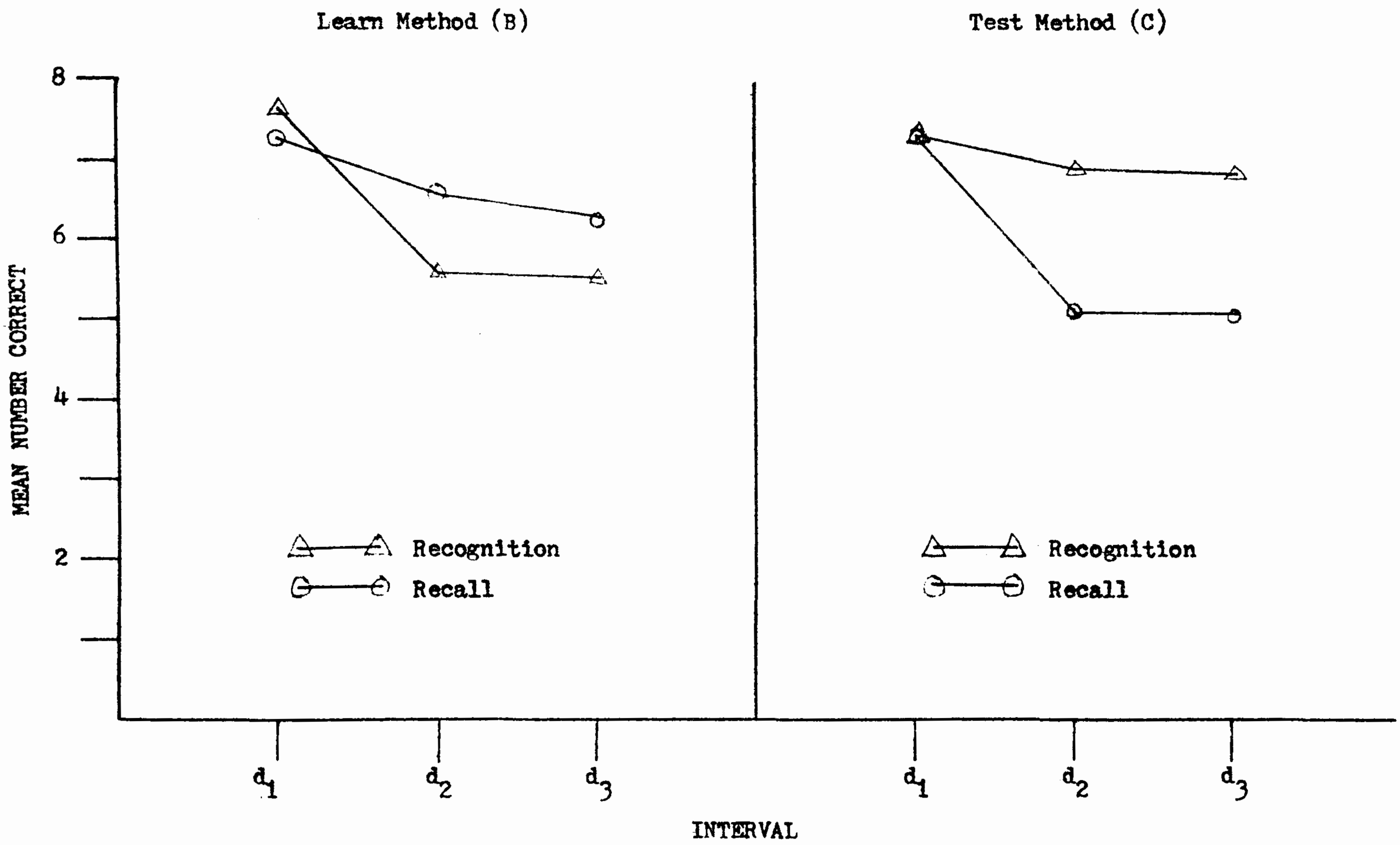

Fig. 2. Profiles of interaction of learn method with interval (BXD) and test method with interval (CXD).

- - - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - - -

reveals no differences between test measures at either method of learning. At the 24 hour interval the profile indicates a departure from parallel as those Ss who learned under recognition and were tested by recall scored significantly poorer than those who were tested by recognition. Those who learned by recall method scored better on the recognition test than those who received the recall test. The same pattern prevailed at the 72 hour interval. Regardless of the method of instruction, higher scores were obtained by recognition tests than by recall tests.

It is important to note that the learning factor (B) was not significant, $F < 1$, nor was the learning factor x test factor (BC) interaction significant, $F(1, 140) = 1.16$. The test of the hypothesis is made by comparing recognition and recall results under conditions of equal original learning. In this experiment there are ten comparisons that can be made since there are five levels of learning under two conditions. Table 4 presents a composite of F tests performed on the means. In all instances but two, the a_1 and a_4 levels under recall learning, the F was large enough to reject the hypothesis of no difference. For all levels of A combined, the

- - - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - - -

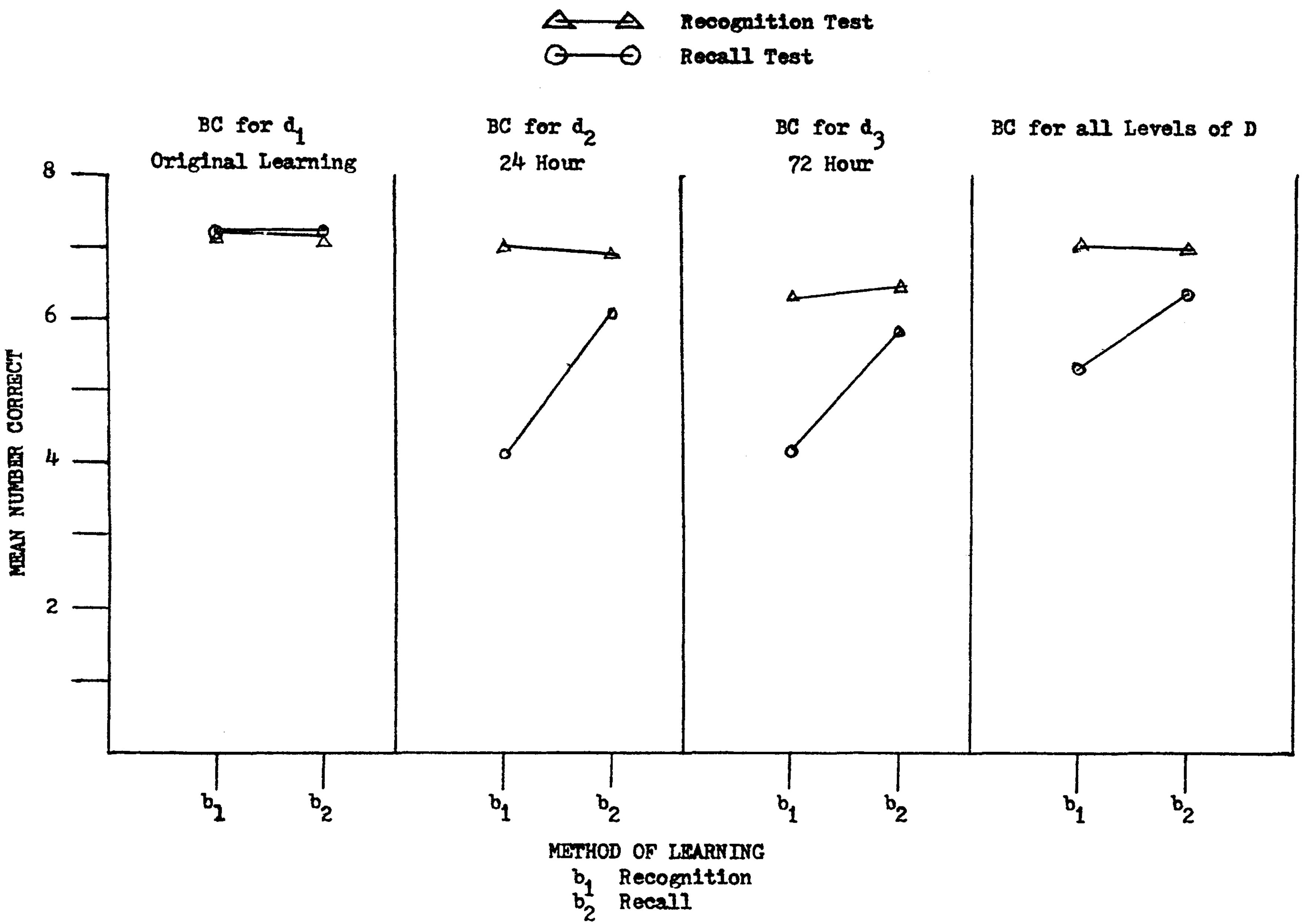

Fig. 3. Profile for interaction of learning method x test method x interval (BCD).

TABLE 4

Value for F When Recognition and Recall Scores for Retention are Compared

	CATEGORY				
	a_1	a_2	a_3	a_4	a_5
RECOGNITION	12.18*	19.54*	71.23*	36.00*	50.27*
RECALL	.13	3.65*	5.15*	.85	33.27*

*p <.05

comparisons between recognition and recall were significant, $F(1, 240) = 34.46$ $p < .01$ for recognition learning and $F(1, 240) = 5.15$ $p < .05$ for recall learning.

DISCUSSION

The results of this experiment support Postman and Rau (1957) and their conclusion that "tests of recognition yield higher scores than do tests of recall." Bahrick's contention that overlearning and easy recognition tests accounts for most of the differences between tests is undoubtedly valid for experiments conducted prior to 1964; however, in this experiment significant differences were obtained even among levels of overlearning. The overlearning effect documented by Postman (1962) was evident. Retention, however, as measured by number of items retained on both tests, increased at a negatively accelerated rate. In this respect the retention curve more approximates that of Krueger (1929).

Postman, Jenkins, and Postman (1948) found recognition is poorer after a recall test than before. They also found that recognition tests followed immediately by a recall test tend to facilitate recall. This was not true in this experiment; however, a true comparison cannot be made as 24 hours elapsed between the two tests whereas Postman et al. tested immediately. One indicator of the differences in difficulty may be the wide margin between the recognition scores and recall scores when learning was by the recognition method. Even though Ss were not told of a second testing session they appeared to have developed a set as to what type of re-test they would receive. Several casual comments were made as to the increased difficulty of the recall test. Of equal importance is the "elation effect" found when Ss were tested by recognition after learning by recall. Many Ss voiced approval of the recognition test over the recall measure.

It may be argued that recognition tests do not actually give higher scores; in actuality what may be happening is repression of recall scores. Kintsch (1968) reported that organization facilitated recall but had little effect on recognition. The list of letter-number pairs presented to Ss in this experiment could not be considered amenable to chunking or blocking; however, the performance of those Ss who learned by the recall method was not statistically significant from those Ss who learned by recognition method when measured at the end of original learning. If list organization had been a significant factor there would have been a difference in original learning.

A difference in the cues provided cannot be considered to be a significant factor in this experiment. McNulty's (1965) findings of a significant difference at various levels of English approximations demonstrated the importance of partial cues in the recognition process. In this instance however, the same cues, in the same order for each test, were provided. If memory traces could have been stimulated by partial cues the opportunity was proffered for both recognition and recall tests.

Perhaps the most telling criticism of any experiment testing measures of retention is that the comparison may be between two methods which are qualitatively different. It may well be that recognition and recall cannot be compared directly as an item may be recalled and then recognized as correct or incorrect. Those proponents of the dual process theory, Cofer (1967), Kintsch (1968), provide data that recall involves a retrieval phase whereas recognition does not.

Another line of reasoning leads to support for the dual process concept. Associative interference as alluded to by Postman and Rau (1957) and Postman and Stark (1969) affects recall but not recognition since the items are provided. Therefore, the experimental design must be based on recall or recognition of a single item as opposed to a list which generates interference. Only when associative interference is equated can definitive conclusions be reached.

The problems of how items are stored and retrieved has yet to be solved conclusively. A more definitive answer will be available when a satisfactory model is postulated and experimental conditions quantified. Until that time, the practical answer is that tests of recognition provide higher scores than tests of recall.

References

Anderson, N. S. Word associations to individual letters. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 541-545.

Bahrick, H. P. Retention curves: Facts or artifacts? *Psychological Bulletin*, 1964, 61, 188-194.

Battig, N. F., & Spera, A. J. Rated association values of numbers from 0-100. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 200-202.

Bjork, R. A. Positive Forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 255-268.

Bruce, D., and Fagan, R. More on the recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1970, 85, 153-154.

Butterfield, E. C., Belmont, J. M., and Peltzman, D. J. Effects of recall requirement on acquisition strategy. *Journal of Experimental Psychology*, 1971, 90, 347-348.

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: Wiley, 1967.

Darley, C. F., and Murdock, B. B. Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, 1971, 89, 66-73.

Davis, J. C., and Okada, R. Recognition and recall of positively forgotten items. *Journal of Experimental Psychology*, 1971, 89, 181-186.

Davis, R., Sutherland, N. S., & Judd, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, 61, 422-429.

Deese, J., & Hulse, S. H. *The psychology of learning*. New York: McGraw-Hill, 1967.

Ebbinghaus, H. Memory: *A contribution to experimental psy chology*. New York: Teachers College, Columbia University, 1913.

Estes, W. K., & DaPolito, F. Independent variation of information storage and retrieval processes in paired associate learning. *Journal of Experimental Psychology*, 1967, 75, 18-26.

Grasha, A. F., Reichmann, P., Newman, A., & Fruth, F. Single-trial recall and recognition memory under conditions where the number and availability of responses are equated. *Journal of Experimental Psychology*, 1971, 90, 306-310.

Greenfield, L. Ease of learning and the effect of over-learning on associative strength, 1969. Cited in *Dissertation Abstracts International*, 1969, 30, 2926B-2927B.

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

Krueger, W. C. F. The effect of overlearning on retention. *Journal of Experimental Psychology*, 1928, 12, 71-78.

Lewis, M. Q. Categorized lists and cued recall. Journal of Experimental Psychology, 1971, 87, 129-131.

Loftus, G. R. Comparison of recognition and recall in a continuous memory task. Journal of Experimental Psychology, 1971, 91, 220-226.

Luek, S., McLaughlin, J. P., & Cicala, G. Effects of blocking of input and blocking of retrieval cues on free recall learning. Journal of Experimental Psychology, 1971, 91, 159-161.

Luh, C. W. The conditions of retention. Psychological Monographs, 1922, 31, (3, Whole No. 142).

McNulty, J. A. An analysis of recall and recognition processes in verbal learning. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 430-436.

Postman, L. Retention as a function of degree of overlearning. Science, 1962, 135, 666-667.

Postman, L., Jenkins, W. O., & Postman, D. L. An experimental comparison of active recall and recognition. American Journal of Psychology, 1948, 61, 511-519.

Postman, L., & Rau, L. Retention as a function of the method of measurement. University of California Publications in Psychology, 1957, 8, 217-270.

Postman, L., & Stark, K. The role of response availability in transfer and interference. Journal of Experimental Psychology, 97, 168-177.

Strong, E. K. The effect of time interval upon recognition memory. Psychological Review, 1913, 20, 339-372.

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

Winer, B. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Wood, G. Retrieval cues and accessibility of higher-order memory units in multitrial free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 782-789.

Instructions

APPENDIX A

Original Learning Session:

"This class has been selected to participate in a learning experiment designed to determine the difficulty of selected letter-number pairs. Although your scores will not be considered in computing your course grade you are encouraged to do your best.

At this time please refer to the instruction page on the front of the test booklets that have been passed out and follow along with me. (Hold up booklet, wait until all Ss begin reading.) A list of eight items, composed of single letters and two digit numbers will be projected on the screen in front of you. Your task is to learn the list of items. The list will be presented 5 times. Each item will be projected on the screen for 5 seconds. The word "START" will appear on the screen 5 seconds before the first item. The word "STOP" will be projected at the end of the trial. A test will be administered at the end of each trial. You will have 30 seconds to record your answers in this test booklet. Once you have completed a test page turn the page and do not refer to it again. Do not turn ahead in the booklet until told to do so. Place your name and the date in the space provided at the top of this page. If you have any questions concerning the procedure to be followed please ask them at this time."

At the end of each trial the following instructions were given, "Turn now in your test booklet to the page marked Trial ____ and record your answer in the space provided. You have 30 seconds." After a 30 second interval, "Stop writing! Turn that test page if you have not already done so and watch the screen for the beginning of trial ____."

Instructions for 24 hour and 72 hour sessions:

"Recently this class participated in a learning experiment designed to determine the difficulty of eight letter-number pairs. A 30 second retention test will be administered to determine which items have been retained. When the "start" signal is given, please turn the test sheet over and record your responses in the spaces provided. You will have 30 seconds to complete the test. When you have finished the test or when "stop writing" is announced please print your name and the date at the top of the page in the space provided."

After the 72 hour test each S was asked to indicate on the back of the test sheet if they had participated in a similar learning experiment within the past calendar year. If yes, the S was contacted and interviewed to determine if their prior experience rendered them unusable in this experiment.

APPENDIX B   Sequence of Presentation

	Order 1	Order 2	Order 3
1	H 61	Q 35	K 97
2	N 43	L 86	F 28
3	L 86	R 72	Q 35
4	K 97	W 59	L 86
5	W 59	F 28	N 43
6	Q 35	K 97	W 59
7	R 72	N 43	H 61
8	F 28	H 61	R 72

	Order 4	Order 5
1	Q 35	N 43
2	F 28	Q 35
3	H 61	F 28
4	L 86	H 61
5	W 59	K 97
6	R 72	R 72
7	N 43	W 59
8	K 97	L 86

VITA

David F. Prim was born on August 24, 1940, in Charlotte, North Carolina. He attended public elementary schools in Moore County, North Carolina, and was graduated from Southern Pines High School in 1958. After a year at Florida Southern College he entered the University of Richmond as a psychology major in 1960. In 1962 he was graduated with a Bachelor of Arts degree and immediately entered the Regular Army as a second lieutenant. After a series of tours in Europe, Asia, Turkey, and the United States, he returned to the University of Richmond to enter graduate school in 1971. After receiving his Master of Arts degree, Major Prim will continue his education at the Army Command and General Staff College at Fort Leavenworth, Kansas.