RABENOU, Bijan, 1943-
METHODS OF PRESENTATION IN PAIRED ASSOCIATE LEARNING: A TEST INTERFERENCE AND RETENTION INTERVAL HYPOTHESES.

The University of Oklahoma, Ph.D., 1975
Psychology, experimental

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

METHODS OF PRESENTATION IN PAIRED ASSOCIATE LEARNING: A TEST INTERFERENCE AND RETENTION INTERVAL HYPOTHESES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

BIJAN RABENOU

Norman, Oklahoma

1975

METHODS OF PRESENTATION IN PAIRED ASSOCIATE LEARNING:
A TEST INTERFERENCE AND RETENTION INTERVAL HYPOTHESES

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

I am profoundly in debt to Dr. N. Jack Kanak for his continuous help and guidance throughout the past four years. The devotion he has shown to the science of psychology and to the training of his students, both in and out of the traditional academic setting, is an inspiration so deeply stored in my memory that it is fully immune to interference mechanisms and yet so accessible as to be retrieved with the subtlest of cues. I wish to extend my deep gratitude to Dr. Robert F. Weiss, not only for introducing me to the field of social learning research and theory construction, but also for his consistent social reinforcement and for his serious treatment of my humorous ideas. Many thanks are due to Drs. Larry E. Toothaker and W. Allen Nicewander who have increased my knowledge of theoretical statistics significantly. My thanks are also due to Dr. Roger L. Mellgren for making even the study of rats a very interesting subject.

//GO.SYSIN DD *

TABLE OF CONTENTS

METHODS OF PRESENTATION IN PAIRED ASSOCIATE LEARNING: A TEST INTERFERENCE AND RETENTION INTERVAL HYPOTHESES

INTRODUCTION

The anticipation (ANT) method of presentation has been generally used in investigations of paired associate (PA) learning (Battig, 1965). In this technique the stimulus member of a pair is presented alone in the anticipation period during which the subject is expected to respond with the corresponding response member. The sequence is then followed by the presentation of the stimulus-response pair during the feedback period. It could be assumed that the "immediate knowledge of results," analogous to "reinforcement," associated with the ANT method should result in a more efficient rate of PA learning. However, Battig and Brackett (1961, 1963) demonstrated that PA performance under the ANT method was inferior to performance under the study-test (ST), or recall, procedure in which all stimulus terms are presented one after the other on a test trial, followed by individual presentations of all the stimulus-response pairs on a study trial. Subsequent research has revealed that superiority of the ST method is not a universally obtained phenomenon. Lockhead (1962) found no difference in PA performance under the two methods of presentation and the results of seven experiments investigating the effects of number of related variables on methods of presentation of PA learning prompted Cofer, Diamond, Olson, Stein and Walker (1967) to conclude that the facilitative effect of the ST method is not robust and may be list-specific. However, of 37 studies reviewed by the author, 24 have reported the ST method to be significantly superior to the ANT method, 10 have reported null effects, and only 3 have found significant superiority favoring the ANT method. The position, therefore, taken here is that the ST method is generally superior when differences are found.

Three major theoretical views have been expressed in an attempt to explain the method of presentation effect. Battig and Brackett (1961) assumed the complete tempral separation between the presentation of stimulus-response pairs and the presentation of stimulus terms alone to be the major source of facilitation in PA learning under the ST method. They argued that this temporal separation in effect facilitates the division of the two behavioral processes of producing a previously learned correct response, or "performance," and the learning of a new stimulus-response association, or "learning." In the ANT method both of these processes occur in close temporal contiguity leading to mutual interference and a consequent reduction in learning efficiency.

Kanak and his associates (Kanak & Neuner, 1970; Kanak, Cole & Eckert, 1972; Duffy & Kanak, in press) have extended Battig and Brackett's theory to include the confidence threshold of subjects as a major factor contributing to the outcome of the results. In addition, these authors have substituted the more contemporary terminology of "storage" and "retrieval" for the terms "learning" and "performance." The assumption is made that in the ANT procedure subjects tend to avoid responding with an "available" item more than under the ST method until the subject is relatively more confident of the correctness of his response. The setting of a higher confidence threshold is assumed to be a by-product of the confusion of "competition" stemming from the rapid alternation between storage and retrieval processes in the ANT method. This hypothesis was supported by results reported by Kanak and Neuner (1970) in that the ANT method produced a higher rate of omission errors even though the two methods did not differ in the response learning stage (Underwood and Schulz, 1960) measures. The "temporal separation" concept of Battig and his associates and its combination with the "competition between storage and retrieval processes" conception of Kanak and his associates will hereafter be referred to as the "interference hypotheses."

The third hypothesis is the "retention interval hypothesis" proposed by Izawa (1972). The retention interval hypothesis is concerned with the number of intervening items between the feedback presentation of a given pair (B_n) and its subsequent anticipation presentation (A_n). This interval has a range of 0 to 2n-2 in the ST method and 0 to 4n-4 in the ANT method, where n is the number of pairs in the list. Thus the mean retention interval for a given pair is twice as long in the ANT method (2n-2) as it is in the ST procedure (n-1), assuming random item presentation on both methods.

Izawa assumes that the longer retention interval in the ANT method produces greater losses in short-tem memory and subsequently results in inferior acquisition of a given pair. The distribution curves of retention intervals of ST and ANT methods as a function of the number of intervening events under random item presentation reveals a large overlapping area under the two curves (Izawa, 1972; see Figure 1). Izawa claims that the overlapping area of the two curves can explain the inconsistent results in the literature. However, Battig (1973) has justly called this claim "fallacious." As Battig has argued, individual ANT pairs with overlapping length of retention intervals with individual ST pairs have to be compensated for by other anticipation pairs with longer retention intervals than any ST pairs.

Support for the retention interval hypothesis is reported by Izawa (1972, 1974). Izawa (1972) found that pairs with longer retention intervals produced inferior paired associate performance than pairs with shorter intervals. In addition, a predicted interaction of list length and method of presentation has been reported (Izawa, 1974). No difference between the two methods of presentation was obtained with short lists where the length of the retention intervals in both methods were short enough to fall in the realm of short-term memory or with very long lists in which the retention intervals are beyond the short-term memory scope. But superiority of the ST over the ANT method was observed with lists of medium lengths as expected by the retention interval hypothesis.

The present study was designed to test the retention interval and interference hypotheses by reducing the length of the retention intervals gradually from the ANT to the ST method, while keeping the list length constant. This manipulation was achieved by variable groupings of anticipation and feedback presentations in eight different conditions with a list of 24 pairs. At one extreme, the traditional ANT method, hereafter designated as condition 1, a single anticipation presentation was followed by its corresponding single feedback presentation. At the other extreme, the ST procedure, designated as condition 24, 24 successive anticipation or test presentations were followed by 24 successive feedback or study presenatations. In the other six conditions, the groupings of feedback and anticipation presentations consisted of 2, 3, 4, 6, 8, and 12, in conditions designated by the same numbers. For example, in condition 4, four anticipation presentations were followed by four corresponding feedback presentations, which preceded presentation of another four stimulus terms in the list, followed by four corresponding feedback presentations.

As Izawa has pointed out, the range of retention interval length in condition 1 is from 0 to 4n-4. In the present extension of Izawa's formulation, as the size of groupings increases from 1 to 24, the upper range of the number of intervening events between the feedback presentation of a given pair and its subsequent anticipation presentation, or the length of the retention intervals, decreases by 2Δ, where Δ is equal to the difference between the number of anticipation or feedback presentation groupings of the respective conditions and that of condition 1. The lower value of the range in all conditions remains a constant zero. Assuming a symmetrical distribution of retention intervals for all conditions as a function of the number of intervening events under random item presentation, the mean of the retention interval for each condition can be determined by dividing the upper limit of the range of each condition by two. Thus the mean retention interval length of each condition can be expressed by the general formula of $(2n-2)-\Delta$.

The retention interval hypothesis predicts increasingly better performance as the size of groupings increases from 1 to 24. However, the same prediction is made by the Battig and Brackett and/or Kanak and Neuner hypotheses. As the size of the grouping increases, the separation of storage and retrieval processes also increases, resulting in a decrease of interference potential and consequently superior performance.

The grouping manipulation produces a separation of storage and retrieval, but it also introduces the confounding variable of delay of feedback presentation, or reinforcement. Accordingly, another set of eight conditions were included to serve as a control for delay of reinforcement and provide differential predictions for the interference and the retention interval hypotheses. This set of eight conditions, delayed feedback (DF), had exactly the same presentational orders and groupings of anticipation items as the previously described set of conditions, termed immediate feedback (IF). However, in the DF conditions the feedback presentation of the first 12 pairs of the respective IF conditions were replaced by the last 12 feedback pairs in each presentation of the list and vice versa. This procedure resulted in two types of pairs depending on whether the anticipation presentation occurred in the first or the second half of the list. For half of the pairs (type I) anticipation presentation was placed in the second half of the list and thus feedback presentation preceded anticipation presentation. For the other half (type II), with anticipation presentation in the first hald of the list, feedback presentation followed anticipation presentation. The groupings of particular pairs in the DF conditions corresponded exactly to that of their respective IF conditions.

Conditions DF-24 and IF-24 do not differ in their length of retention intervals and delays of feedback, the number of intervening items between anticipation presentation of an item and its corresponding feedback presentation. However, the lengths of retention intervals and delay of feedback vary considerably in all other conditions of IF and DF. The length of delay of feedback in the IF-1 conditions is always equal to zero. As the size of groupings or items increases in the IF conditions the upper limit of the range of feedback delay length increases by 2Δ while the lower limit of the range remains zero, resulting in a mean increase of Δ. Thus the mean delay of feedback in the IF-24 condition equals to 23 or n-1, a figure presented by Izawa (1972). In the DF conditions, however, the length of feedback delay depends upon the type of paurs. In the DF-1 condition the length of feedback delay for type II pairs has a constant value of n. For type I pairs the length of feedback delay varies from 2, when A_n is at the end of thelist in one presentation and B_n is the first item on the following presentation of the list, to 3n-2. The upper limit occurs when A_n is presented as the first item in the second half of the list in one presentation and B_n is presented as the last item in the following presentation of the list.[1] The mean length of delay of feedback, therefore, is equal to 5n/-4 in the DF-1 condition. Thus the delay interval of DF-1 condition is longer than that on the IF-24 condition which itself has the longest delay interval among the IF conditions.

The length of the delay interval increases with the size of groupings from 1 in DF-1 to 12 in DF-12. The increase for type II pairs ranges from zero to 2Δ, with a mean of Δ. For type I pairs the lower limit of the range of the delay interval increases by 2Δ, while the upper limit of the range remains constant across all conditions. The mean increase in the length of delay intervals, thus, equals to Δ. While increases in delay intervals in corresponding IF and DF conditions are equal, it should be remembered that the absolute delay interval in each DF condition exceeds that of its respective IF condition by 5n/4 events.

An opposite trend is operative in the retention interval lengths of the IF and DF conditions. In the DF-1 condition the lengths of the retention intervals of type I pairs have a constant value of n-2, since feedback presentation precedes the anticipation presentation for these pairs.

---

[1] In cases where two A_ns followed each other without an intervening B_n presentation, the shorter delay interval was considered in these calculations. The same procedure was used in calculations of retention interval lengths.

For type II pairs the length of the retnetion interval has a range of zero, when B_n is the last item in one presentation and A_n is the first item on the next presentation of the list, to 3n-4, when B_n is the first feedback presentation in one list presentation and A_n is the last anticipation item in the first half of the list in the next presentation. The resultant mean retention interval length is equal to (5n-8)/4. This value is much lower than the mean retention interval length of the IF-1 condition and only slightly higher than that for IF-24. As the size of groupings increases from 1 in DF-1 to 12 in DF-12 the length of the retention interval decreases. The decrease for both types of pairs ranges from zero to 2Δ with a mean decrease of Δ. Thus the decrease of retention interval length with increases in grouping size is identical in the IF and DF conditions while the baseline mean retention interval length (for IF-1 or DF-1) has a much smaller value in the DF conditions.

Table 1 presents the mean retention and delay interval lengths for each of the 16 conditions in terms of n and Δ as well as the actual number of intervening events for the list length and the size of groupings employed in the present experiment. The retention interval lengths in the DF conditions are substantially shorter than those in IF conditions. Thus the retention interval hypothesis predicts superior paired associate performance under the DF conditions. Moreover, as the size of groupings increases, the retention interval length decreases throughout all IF conditions and up to condition 12 in the DF conditions. The retention interval hypothesis, therefore, further predicts a gradual improvement in performance as the size of groupings increases. As Table 1 shows, the lengths of the retention interval for conditions DF-8 and DF-12 are shorter than those of condition IF-24 and DF-24. Paired associate performance is thus expected to be best in those two conditions.

Somewhat different predictions are made by the interference hypotheses. The interference hypotheses share with the retention interval hypothesis the prediction that increases in size of grouping results in superior performance. However, the prediction of the interference hypotheses with regard to comparisons of DF and IF conditions is opposite that of the retnetion interval hypothesis. This prediction stems from the interference effect of retrieving a given item and storing another item. This effect added to the competition or interference casued by the rapid alternation of storage and retrieval processes is expected to produce the poorest performance in condition DF-1. As the size of groupings increases, however, the inferiority of DF to IF conditions is expected to decrease since the two conditions become more uniform.

Method

Design. The design of the experiment constitutes a 2 x 8 factorial with two levels of delay of feedback (IF & DF) and eight methods of presentation based on the size of grouping of anticipation or feedback presentations (1, 2, 3, 4, 6, 8, 12 & 24), resulting in sixteen independent conditions.

Subjects. One hundred-sixty undergraduate students enrolled in introductory psychology classes at the University of Oklahoma participated in the experiment as a part of an option among requirements of the course. All subjects were naive to PA learning. Each subject was randomly assigned to one of the 16 treatment conditions (n = 10) upon appearance at a laboratory within the restriction of achieving equal cell n's.

Materials. The materials selected for list construction were similar to those employed by Izawa (1972). Two lists of 24 pairs each were constructed for the purpose of generality. The stimulus terms were CVCs with associative values ranging from 1.50 to 1.69 with means of 1.56 and 1.57, based on the Noble (1961) norms, for the two lists. The stimulus terms within each list never shared more than one letter. The responses consisted of two-digit numbers from 23 to 96. The following restrictions were placed on selection of response terms for each list. No number ending with zero was included. Numbers with repeating digits (e.g., 33) were excluded. No two numbers involving a simple reversal of digits were used in the same list. For example, if 35 was included as a response term in one list, 53 was not selected for the same list. The numbers had associative values ranging from 0.70 to 2.07 with means of 1.32 and 1.33, based on the Battig and Spera (1962) norms, for the two lists. Each stimulus was randomly paired with a response. Half of the subjects in each condition learned one list and the other half learned the other list.

Four random serial orders of presentation of the list were employed in each condition. The serial order of anticipation presentations in all the conditions were exactly the same for each order of presentation. In the IF conditions the groupsings of anticipation and feedback presentations was varied in eight different conditions. The feedback pairs with each grouping corresponded to the stimuli presented in the immediately preceding group of anticipation presentations, but the order of feedback presentation within each grouping was randomly selected; illustratively for IF-4: A_1, A_2, A_3, A_4, B_3, B_2. B_4, B_1, A_5, A_6, . . . B_{24}, B_{23}. In the DF conditions the feedback for the first 12 anticipation presentations of the respective IF conditions were replaced with the feedback for stimulus terms of the last 12 presentation, and vice versa, in each presentation of the list; i.e., in DF-1 condition: A_1, B_{13}, A_2, B_{14}, A_3, B_{15}. . . A_{13}, B_1, A_{14}, B_2, . . . A_{24}, B_{12}. For eight of the 24 pairs the anticipation presentation was placed in the first half of the list in three randomized orders of the list and in the last half in one randomized order. For another set of eight pairs the anticipation presentation was in the first half of the list in two and in the last half of the list in the other two randomized orders of presentation, while for the remaining eight pairs anticipation presentation was placed in the last half of the list in three orders and in the first hald in one presentational order of the list.

Procedures. The lists were presented on a Lafayette memory drum at the rate of 2:2 sec. The intertrial interval for all conditions except IF-24 and DF-24 was 4 sec. In DF-24 and IF-24 conditions two 2 sec intertrial intervals separated the study and test trials. Subjects were given standard paired associate task instructions and were further informed about the nature of the groupings of the items and of immediate or delayed feedback. In addition, an example of the presentational method involved was shown to the subject who was explicitly informed that those items would not appear in the list. All subjects were presented one study trial in which all 24 pairs appeared at a 2 sec rate, prior to onset of the practice trails. The interval between the initial study trial and the first practice trial corresponded with the intertrial interval. Subjects had 24 practice trials unless they reached the criterion of one perfect trial in which case practice was terminated. Subjects who did not reach the *a priori* determined criterion of eight correct responses in a single trial within the 24 practice trials were not included in the analysis of the data and were replaced by the subject appearing next in the laboratory. A total of 18 subjects did not reach the criterion. Of these, 3 were assigned to DF-8 condition; 2 each in conditions, DF-4, DF-3, IF-3, IF-2, IF-1, DF-1, and one in each condition of DF-6, IF-6 and IF-8.

Results

Analysis of total errors on the two types of lists employed failed to produce any significant main effects or interactions due to list differences (all $ps > .10$). Accordingly for further analyses the list factor was omitted.

Trials to criterion. Analysis of the number of trials to reach the criterion of 8 correct responses in a single trial produced a significant main effect of method of presentation, F (7, 144) = 3.01, $p < .01$, and a nonsignificant main effect of delay of feedback and interaction with methods of presentation (both $Fs < 1$). The mean number of trials to criterion for conditions 1, 2, 3, 4, 6, 8, 12 and 24 were, respectively, 13.55, 13.40, 14.65, 13.55, 13.25, 12.85, 10.80, and 8.15. Tukey's multiple comparisons among the means revealed significant superiority of performance in the conditions 24 to conditions 1, 2, 3, and 4 (all $ps < .05$). The difference between conditions 6 and 24 only approached a significant level ($p < .10$), while no significant differences were obtained among any other conditions. The analysis of trials to criterion of 4 correct responses produced similar, though less potent, results. The superiority of condition 24 was significant only in comparison with condition 3.

Error measures. Analysis of total number of errors corresponded with the results of the analyses of trials to the criterion of 8 correct responses. Only the main effect of method of presentation reached significance, F (7, 144) = 3.37, $p < .005$. Although delay of feedback produced more errors, the main effect of feedback fell short of significance ($p > .10$) as did the interaction of the two variables ($F < 1$). The mean total errors for conditions 1, 2, 3, 4, 6, 8, 12, and 24 were, respectively, 398.00, 367.50, 406.05, 400.30, 494.40, 368.80, 351.70, and 297.00. Tucky's multiple comparisons resulted in significant superiority of conditions 24 to conditions 1, 3, 4 and 6 (all $\underline{p}$s $< .05$) with all other comparisons nonsignificant. The relatively superior performance of condition 2 was unexpected. The effect may be due to two subjects in condition IF-2 who learned the list at a very fast rate and reached the criterion of one perfect trial within 12 trials. This rate of learning is unusually fast considering that only 13 other subject in all conditions reached this criterion within twenty-four trials and of those only one other was in all of conditions 1, 2, 3, and 4 of IF or DF combined. When the data of those two subjects were excluded, the mean number of errors for condition 2 was 392.77, more in line with the mean of adjacent conditions 1 or 3. It seems, therefore, that the superiority of condition 2 can be accounted for by sampling error.

Error measures were also analyzed in terms of total errors over blocks of four trials and proportions of errors in blocks of four trials. The proportions of errors were calculated by dividing the total number of errors in each four trials by 96, a multiplicatory product of four trials and the number of items per trial. Since both dependent variables produced identical results, only the errors per blocks of trials analysis will be reported here. A 2 (feedback delay) by 8 (methods of presenation) by 6 (blocks of trials) analysis of variance produced significant main effects of methods of presentation, $\underline{F}$ (7, 144) = 3.33, $\underline{p} < .005$ and blocks of trials, $\underline{F}$ (5, 720) = 914.65, $\underline{p} < .001$. More importantly, the interactions of blocks of trials with delay of feedback, $\underline{F}$ (5, 720) = 4.61, $\underline{p} < .0005$, and with methods of presenation, $\underline{F}$ (35, 720) = 2.45, $\underline{p} < .0001$, were both highly significant. The interaction of blocks of trials and delay of feedback revealed an increasing superiority of the IF conditions over DF conditions as practice progressed. The differences between the two conditions were not reliable in the first four blocks but reached the usually accepted level of significance in the last two blocks (Tukey's comparison; both $ps < .05$).

Table 2 presents the means and standard deviations of errors for each of method of presentation conditions at each of the six blocks of trials. It is evident from the table that the differentiation among the methods increased with practice. Tukey's multiple comparisons verify this impression. No reliable differences in the conditions were obtained in the first two blocks of trials. Condition 24 produced significantly fewer errors than conditions 1, 3, 4, and 6 in the next two blocks and was significantly superior to all conditions except 12 in the last two blocks of trials (all $p < .05$). Although condition 12 consistently produced fewer errors across all blocks than all other conditions with groupings of smaller size, the differences between condition 12 and these other conditions were relatively small and nonreliable.

The error measures were further divided into response intrusion (intra- and extralist combined) and response omission errors to examine the possible differential interaction of these measures with the independent variables. Analysis of variance on intrusions per block of trials produced a significant main effect of blocks F (5, 720) = 11.51, $p < .0001$ and its interaction with methods of presentation, F (35, 720) = 1.89, $p < .002$. The main effect of method of presentation and the interaction of feedback delay with blocks, significant in total errors analysis, were not reliable in this analysis (both $ps > .15$). Significantly fewer intrusion errors were made in the last block of trials than in all other blocks except the fifth (all ps $< .05$). Moreover, intrusion rates in the second, third, and fourth blocks were higher, though not significantly so, than all the remaining three blocks. More interesting results are involved in the interaction of blocks with methods of presentation. Tukey's multiple comparisons between the first and the last block for each method revealed no reliable difference in conditions 1, 2, 3, 4 and 6. However, intrusion errors were significantly lower in the last block than in the first block for conditions 8, 12 and 24.

Intrusion errors, however, may not be a very sensitive measure, as the number of intrusion errors is necessarily limited by the total number of errors. The proportion of intrusion errors to total errors may provide a more sensitive dependent variable. The proportions of intrusion errors were obtained by dividing the number of intrusion errors by the total number of errors (intrusions plus omissions) in each block for each subject. Analysis of variance on this measure resulted in significant main effects of feedback delay, $F\ (1, 144) = 3.92$, $p < .05$, and blocks, $F\ (5, 720) = 40.46$, $p < .0001$, as well as an interaction of blocks with methods of presentation, $F\ (35, 720) = 1.50$, $p < .02$. The proportion of intrusion errors was greater in IF conditions than in DF conditions and increased with practice. Table 3 presents the means and standard deviations of each method at each of the six blocks of trials. As can be seen from Table 3, the proportion of intrusion errors increased in all conditions as practice progressed. However, the degree of increase was much larger in conditions with small groupings than those with larger groupings as was confirmed in the post hoc analyses. Comparisons of the first and last blocks revealed significant increases in intrusion errors in conditions 1, 2, 3, 4, and 6 (all ps < .01) while the difference between the first and last blocks in the remaining three conditions (8, 12 & 24) were not reliable (all ps > .05). These results, combined with those obtained in the analysis of number of intrusion errors per block of trials, indicate that while intrusion errors decreased with decreases in total errors in conditions 8, 12 and 24, the rate or proportion of intrusion errors remained more or less constant throughout practice in the remaining conditions.

Moreover, the lack of interaction of feedback delay and blocks of trials in the analysis of intrusion errors suggests that such an interaction on the proportion measure is mainly due to omission errors. The analysis of omission errors provided direct evidence for this notion. The number of omission errors per block and the proportion of omission errors produced similar results. For the sake of brevity only the analysis of the latter measure is reported here. Analysis of the proportion of omission errors per block of trials produced significant main effects of delay of feedback, F (1, 144) = 5.70, $p < .02$, and blocks F (5,720) = 61.48, $p < .0001$, and a significant interaction of the two variables F (5, 720) = 2.47, $p < .05$. In addition, the main effect of method of presentation approached the accepted level of significance ($.05 < p < .10$). The proportions of omission errors were greater in the DF conditions than in the IF conditions and decreased as practice trials increased. Moreover, the differences between the DF and IF conditions increased with practice trials. Differences between the two conditions were not significant in the first three blocks of trials, but reached a reliable level in the last three blocks (all ps < .05). The interaction of methods of presentations with blocks, significant in the analysis of total errors, was not reliable in the proportion of omission errors analysis ($F < 1$). This indicates a reduction in omission errors as practice progressed in all conditions of methods of presentation. The significant interaction of methods of presentation with blocks observed in total errors by blocks analysis apparently was caused by intrusion errors.

Number of correct responses. Analysis of data on total number of correct responses and the number of correct responses per block of trials produced results identical to their corresponding total error measures as might be expected. An analysis was also performed on the number of correct responses on the last practice trial for each subject. The IF conditions produced more correct responses on the last practice trial than the DF conditions as indicated by significant main effect of feedback, F (1, 144) = 3.85, p .05. In addition, the main effect of methods of presentation was highly significant, F (7, 144) = 3.48, p .005, indicating the superiority of condition 24 to all other conditions except 12 (all $ps < .05$).

Discussion

Manipulations of methods of presentation produced a significant main effect and/or interacted with blocks of trials in the analysis of all measures except omission errors. The effect was mainly due to the superiority of condition 24 to all other conditions except condition 12. The superiority of condition 24 over condition 1 replicated the finding of previous investigations in which similar materials were employed (Battig & Brackett, 1961; Izawa, 1972, 1974). Increases in the size of the grouping of anticipation and feedback presentations did not produce a potent facilitative effect as predicted from all three major hypotheses concerning method of presentation effects. The lack of differences in performance under conditions 1, 2, 3, 4, and 6 can be readily explained by both the interference and retention interval hypotheses. The small size of groupings employed in these conditions required rapid alternation of retrieval and storage mechanisms. Even in condition 6 the alternation of processes occurred every 12 sec and thus may not have allowed enough time for effective separation of storage and retrieval mechanisms. Moreover, the reduction of the number of intervening events in these groups, specifically under the IF condition, were relatively small and the retention interval may have been long enough to produce substantial short-term memory losses. More puzzling is the performance under conditions 8 and 12. These conditions did not produce reliably better performance than conditions with smaller size groupings. Nonetheless, as Table 2 reveals, condition 12 produced fewer errors in each block of trails than all conditions except 24, and performance under condition 8 surpassed all other conditions except 2, 12 and 24. As mentioned earlier, the unexpected superiority of condition IF-2 to its adjacent conditions may well be an artifact of sampling error. This notion is strengthened by the results of intrusion errors analysis. The proportion of intrusion errors increased significantly as practice progressed in condition 2 as it did in conditions 1, 3, 4, and 6, but unlike conditions 8, 12 and 24 in which little increase in intrusion errors across trials was observed. This result suggests a qualitative difference between the two sets of conditions. The fact remains that condition 24 was never found to be significantly superior to condition 12 and only in the last two blocks of trials did condition 24 show reliably better performance than condition 8, indicating at least some facilitation effect in conditions 8 and 12. Three reasons may be proposed for the lack of reliable superiority of conditions 8 and 12 over conditions with smaller size groupings. First, the size of groupings may not have been large enough to prevent interference between storage and retrieval processes. This seems unlikely since Battig and Brackett (1961) reported superiority of ST over ANT method employing a 12-pair list and Wright (1967) found the same result with an 8-pair list. In both investigations CVC-two-digit pairs were used as on the present study. Second, in condition 24 the separation of anticipation and feedback presentations were accentuated by an intertrial interval of 2 sec, while no such interval was provided in the other conditions. Although all conditions had a 4 sec intertrial interval, the intertrial interval was presented as a means to separate presentation of trials and not feedback and anticipation groupings. It is reasonable to assume that the additional intertrial interval in condition 24 provided an additional cue to prepare the subject to retrieve or store the materials, thus improving the efficiency of these processes. Third, the lack of reliable supremacy of conditions 8 and 12 to conditions with smaller size groupings may have been caused by the low degree of learning achieves in the present experiment. Only 15 subjects (less than 10% of all subjects) reached a criterion of one errorless trial within the allotted 24 practice trials. The error measures indicated that the separation among conditions increased with practice trials. It seems likely that had practice been carried out to the criterion of one errorless trial for all subjects, more reliable differentiation of performance under the different conditions of method of presentation would have been obtained.

The relative improvement of performance in conditions 8, 12 and 24 supports all three major hypotheses. The superiority of DF conditions over IF condition, however, predicted by the retention interval hypothesis, was not obtained. As can be seen in Table 1, the length of retention intervals in the DF conditions are substantially shorter than those in IF conditions. The number of intervening events in condition DF-6 is equal to those in the condition 24 and conditions DF-8 and DF-12 have shorter retention intervals than both DF-24 and IF-24. Thus the results obtained in the present study are indeed opposite the expectations of the retention interval hypothesis. The DF conditions produced more total errors and omissions errors and less correct responses in the last three blocks of trials and also resulted in less correct responses on the last practice trial than the IF conditions. These findings, in part, confirm Lockhead's (1962) finding that a condition similar to DF-1 was significantly inferior to ANT procedure. However, unlike the present study, no difference between the ST and ANT procedures was obtained by Lockhead. The inferior performance of subjects under the DF conditions clearly demonstrates that the length of retention interval cannot be responsible for the superiority of conditions 12 and 24 to all other conditions and cannot explain the differential effect of methods of presentation.

The analysis of intrusion errors revealed that the effect of methods of presentation was caused mainly by a relative increase of intrusion errors in conditions 1, 2, 3, 4, and 6 with practice trials. This finding is in contrast with those reported by Kanak and Neuner (1970) that the ANT method differed from ST mainly in the omission error rate. The contrasting result may stem from differences in materials employed in the two investigations. Kanak and Neuner used highly familiar words. It is possible that the limited nature of the pool of two-digit numbers employed in the present experiment encouraged subjects to "guess" relatively more frequently, resulting in more intrusion errors. Moreover, the length of list used in the present study necessitated each three responses to have the same first digit thus increasing response competition. The interference produced by response competition added to the interfering effect of storage and retrieval mechanism may be responsible for the greater number of intrusion errors emitted by subjects in conditions with smaller size grouping. Although the DF condition produced lower proportions of intrusion errors than the IF conditions, no reliable difference between the two conditions was found in the absolute number of intrusion errors. The lower proportion of intrusion errors in the DF condition apparently was caused by the greater number of total errors in those conditions.

The inferior performance of subjects in DF conditions cannot be explained by delay of feedback. If delay of feedback presentation was responsible for the inferior performance of DF conditions, one would expect poorer performance in conditions with larger grouping size. Exactly the opposite result was obtained. The inferiority of DF conditions seems to be caused by increased interference of storage and retrieval of different items as predicted by the interference hypotheses.

It can be concluded that the results of the present experiment provide substantial support for the interference hypothesis and present strong evidence against the retention interval hypothesis of Izawa.

REFERENCES

Battig, W. F. Procedural problems in paired-associate learning research. Psychonomic Monograph Supplement, 1965, No. 1.

Battig, W. F. Evidence that recall-anticipation differences increase with increasing time interval. Journal of Experimental Psychology, 1973, 100, 399-402.

Battig, W. F. And Brackett, H. R. Comparisons of anticipation and recall methods of paired-associate learning. Psychological Reports, 1961, 9, 59-65.

Battig, W. F. and Brackett, H. R. Transfer from verbal discrimination to paired-associate learning: II effects of intralist similarity, method and percent occurrence of response members. Journal of Experimental Psychology, 1963, 65, 507-514.

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

Cofer, N. C., Diamond, F., Olsen, R.A., Stein, J.S., and Walker, H. Comparison of anticipation and recall methods in paired-associate learning. Journal of Experimental Psychology, 1967, 75, 545-558.

Duffy, J. C. and Kanak, N. J. Confidence thresholds in two methods of paired associate learning as a function of ego strength. American Journal of Psyhchology, in press.

Izawa, C. Retention interval hypothesis and evidence for its basic assumptions. Journal of Experimental Psychology, 1972, 96, 17-24.

Izawa, C. Retention interval hypothesis and list lengths: Comparison of anticipation and study-test procedures. Canadian Journal of Psychology, 1974, 28, 214-224.

Kanak, N. J., Cole, L.E. and Eckert, E. Implicit associative responses in verbal discrimination acquisition. Journal of Experimental Psychology, 1972, 93, 309-319.

Kanak, N. J. and Neuner, S. D. Associative symmetry and item availability as a function of five methods of paired associate acquisition. Journal of Experimental Psychology, 1970, 86, 288-295.

Lockhead, G. R. Method of presenting paired associates. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 62-65.

Noble, C. E. Measurements of association value (a), rated associations (a') and Scaled meaningfulness (m') for the 2100 CVC combinations of English alphabet. Psychological Reports, 1961, 8, 487-521.

Underwood, B. J. and Schulz, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.

Wright, J. H. Effect of stimulus meaningfulness, method of presentation and list design on the learning of paired associates. Journal of Experimental Psychology, 1967, 73, 72-75.

TABLE 1

Mean retention and delay interval lengths for each of the 16 conditions in terms of n and Δ and actual number of intervening events.

Condition	Size of Δ	IF: Means in terms of n and Δ		IF: Means actual number of intervening events		DF: Means in terms of n and Δ		DF: Means actual number of intervening events	
		Retention Interval	Delay Interval	Retention Interval	Delay Interval	Retention Interval	Delay Interval	Retention Interval	Delay Interval
1	-	2n-2	0	46	0	$\frac{(5n-8)}{4}$	$\frac{5n}{4}$	28	30
2	1	(2n-2)-Δ	Δ	45	1	(5n-8)/4-Δ	$\frac{5n}{4}+\Delta$	27	31
3	2	(2n-2)-Δ	Δ	44	2	$\frac{(5n-8)}{4}-\Delta$	$\frac{5n}{4}+\Delta$	26	32
4	3	(2n-2)-Δ	Δ	43	3	$\frac{(5n-8)}{4}-\Delta$	$\frac{5n}{4}+\Delta$	25	33
6	5	(2n-2)-Δ	Δ	41	5	$\frac{(5n-8)}{4}-\Delta$	$\frac{5n}{4}+\Delta$	23	35
8	7	(2n-2)-Δ	Δ	39	7	$\frac{(5n-8)}{4}-\Delta$	$\frac{5n}{4}+\Delta$	21	37
12	11	(2n-2)-Δ	Δ	35	11	$\frac{(5n-8)}{4}-\Delta$	$\frac{5n}{4}+\Delta$	17	41
24	23	(2n-2)-Δ or n-1	Δ or n-1	23	23	(2n-2)-Δ or n-1	Δ or n-1	23	23

TABLE 2

Mean and standard division of errors under each method of presentation at each of 6 blocks of trials.

Condition		1	2	3	4	5	6
1	$\overline{X}$ SD	91.90 4.15	82.05 8.14	69.90 10.45	60.50 12.76	50.35 16.13	42.90 19.64
2	$\overline{X}$ SD	90.45 5.25	77.55 14.61	63.85 21.19	53.70 24.60	44.80 22.23	35.15 20.62
3	$\overline{X}$ SD	92.40 2.57	83.05 6.68	71.85 10.56	62.20 14.66	52.55 17.21	43.85 17.58
4	$\overline{X}$ SD	90.85 4.17	81.50 9.43	70.80 12.55	61.00 15.64	51.25 17.46	42.70 21.63
6	$\overline{X}$ SD	89.75 4.66	78.60 10.70	69.85 16.30	59.70 16.94	52.10 20.20	44.40 19.77
8	$\overline{X}$ SD	89.50 4.17	77.05 13.08	65.20 18.99	54.65 21.68	46.70 23.82	39.50 23.45
12	$\overline{X}$ SD	86.80 6.41	73.15 12.47	61.55 18.14	52.45 20.15	42.00 20.58	35.45 21.57
24	$\overline{X}$ SD	88.55 4.94	68.80 15.37	52.20 18.11	39.00 18.96	28.10 16.84	20.35 15.75

TABLE 3

Mean and standard division of proportion of intrusion under each method of presentation at each block of trials.

Condition		Blocks					
		1	2	3	4	5	6
1	$\overline{X}$	.17	.24	.31	.31	.37	.42
	SD	.19	.23	.26	.22	.26	.25
2	$\overline{X}$	.21	.28	.32	.30	.34	.37
	SD	.29	.29	.31	.31	.30	.30
3	$\overline{X}$	.09	.14	.19	.22	.25	.25
	SD	.09	.08	.11	.12	.17	.18
4	$\overline{X}$	.11	.14	.16	.19	.25	.29
	SD	.16	.14	.12	.16	.18	.19
6	$\overline{X}$	.09	.12	.18	.22	.21	.25
	SD	.12	.13	.18	.20	.20	.22
8	$\overline{X}$	.17	.20	.22	.24	.23	.22
	SD	.22	.22	.20	.22	.23	.23
12	$\overline{X}$	.16	.20	.23	.23	.25	.24
	SD	.16	.19	.18	.21	.21	.21
24	$\overline{X}$	.14	.20	.24	.24	.23	.22
	SD	.11	.12	.16	.20	.25	.23

APPENDIX A

PROSPECTUS

PROSPECTUS

The anticipation method of presentation (MP) has been generally used in investigations of verbal learning, specifically in studies dealing with paired-associate (PAL), serial (SL) and verbal discrimination (VDL) learning. Battig (1965) reported the number of articles published in three major psychological journals during 1959-1963 in which the anticipation (ANT) method was employed in PAL to be three times as frequent as those using a non-anticipation procedure. Historically, however, another method was more commonly used than the ANT procedure up to the year 1938. In fact, Pennington and Waters (1938), in a study which compared anticipation with a procedure which is now labeled the study-test (ST) or recall method, referred to the latter procedure as the "traditional method" for PAL and introduced the ANT method as a new method of presentation previously used only with SL.

The reason for the recent popularity of the ANT procedure is not immediately clear. One possible reason may be the close correspondence between the ANT method and typical animal learning paradigms. As in animal learning procedures, the anticipation MP in paired-associate learning consists of presentation of a stimulus alone for a period of time during which S emits a response followed by presentation of the stimulus and response paired together analogous to reinforcement in the animal learning methods. Unlike the animal learning model, however, the reinforcement is presented regardless of the correctness of the response in ANT procedure. The ST procedure differs from the ANT method in that it consists of alternating trials of presentation of all the pairs in the list (reinforcement or study interval) and all the stimuli (test interval) in the list. Besides this obvious discrepancy of the ST method with a typical animal learning trial, the ST procedure results in a delay of reinforcement between test and study trials, which may have contributed to its relative lack of use in verbal learning research since the delay variable has been so important historically in animal conditioning procedures.

Although the comparison of the two methods of presentation was reported as early as 1938 by Pennington and Walter, and superiority of performance with the ST procedure over the ANT method was demonstrated by them, interest in systematic comparison of the ANT method with other procedures was not stimulated until the more recent re-introduction of the ST method by Battig and Brackett (1961) and the introduction of the prompting (PRO) procedure by Cook and Kendler (1956).

In the prompting method a stimulus-response pair is presented first followed by presentation of a stimulus alone for a period of time during which the subject is required to pronounce either overtly or covertly the response term they had seen. The test of learning is made on separate trials in which all stimulus members in the list are presented alone and the subject is instructed to respond to each of the stimuli. Both studies found the ANT performance to be inferior to the other methods. These two reports stimulated a large number of investigations examining the effect of MP on performance on verbal learning tasks. Although a review of studies concerning the PRO procedure has been previously reported (Aiken and Lau, 1967), no such review of investigations involving comparisons of the ANT and ST methods of presentation has been attempted. The present paper surveys such investigations. In addition, studies examining the prompting method are included for two major reasons. The first is to give the reader a better perspective of the literature pertaining to methods of presentation. The second, and more important, reason is the similarity between the ST and PRO methods of presentation. For all practical purposes, in both methods learning and performance are temporally separated while, in comparison, learning and performance are interrelated in the ANT procedure.

Battig (1965, 1969) has pointed out a number of disadvantages of the ANT method, the most important being inseparability of measures of learning and performance. Battig argues that the essential property of the ANT procedure with a PA task is that each attempted anticipatory (performance) response is followed by presentation of the correct response and an attempted learning response. Thus, the experimenter's measure of performance in any given trial is contaminated by the immediate occurrence of additional learning, which itself cannot be reflected in the subject's performance until the next trial. Battig further suggests that this simultaneous occurrence of learning and performance is a source of confusion to subjects and cites the reported faster rate of learning under the ST method as evidence of such a confusion. The inherent problem of inseparability of learning and performance in the ANT method is only intensified with a SL task, where the subject is required to both retrieve an item from memory (perform) and store another item in the memory (learn) with every presentation of an item. Battig also argues that individual pairs in the list are almost invariably learned at different rates and to different terminal levels with the commonly used list based criteria. Experimenters have no control over individual pair learning and no effective method for measurement of degrees of learning for each individual pair. Moreover, the ANT method of PAL does not allow measurement of the associative relation of interpair items. Battig claims that the assumption that the primary interest in PAL is examination of intrapair associations is erroneous and feels the influence of the context of other pairs on associative connection of stimulus-response terms is an important aspect of PAL which is usually ignored.

Clearly the first advantage of ANT method, that of learning and performance undetachment can be removed by employment of the ST procedure. As Battig (1969) has reasoned in his defense of the ST method, the separation of study and test trials frees subject's performance from contamination by new learning and allows more accurate measurement of performance than is possible with the ANT method. Such a separation also reduces the subject's confusion. However, Battig's other criticisms of the ANT method can also be applied to the ST method of presentation. The ST method is equally as vulnerable to the problem of unequal learning of individual items and insensitivity to interitem relationship measurements. Although Battig proposed procedural modifications to correct these disadvantages, the discussion of such procedures is well beyond the scope and purpose of the present paper and thus is ommitted here.

Another advantage in employment of ST method lies in its close correspondence with extra-experimental learning situations in adult humans. It is unlikely that anyone trying to store new information into his memory would do so by storing a small chunk of information and then testing himself on that chunk before endeavoring memorization of a new chunk. No student preparing for an upcoming exam ordinarily would try reading on a single concept and testing himself on that concept before going on to the next paragraph. He is more likely inclined to read all the material required before testing his memory and then focusing on re-reading of the unlearned material. This alternation of learning and testing procedures more closely resembles the ST method of presentation than ANT. In learning experiences at earlier stages of development, however, parents and teachers usually require a single response from children and provide immediate feedback. Such a learning experience more closely approximates the anticipation learning situation. The laboratory examination of the comparative efficiency of the two methods in learning with adults and children, therefore, has practical as well as theoretical implications.

An obvious disadvantage of the ST procedure is the delay of feedback, reinforcement, or knowledge of results. The duration of the feedback presentation for a given pair in the ST procedure depends on the serial position of stimulus in the test trial and the serial position of stimulus-response pair on the subsequent study trial, as well as the length of the list. However, regardless of the position of the item and list length, delay of feedback presentation is longer in ST procedure than in the ANT method. Since feedback in a verbal learning task is analogous to reinforcement in animal learning models, it could be assumed that delay of feedback in a verbal learning task has an adverse effect on performance. Empirical data, however, have not substantiated this assumption, as the review of studies on examining delay of feedback in the next section demonstrates.

Delay of Feedback

Numerous experiments have examined the effect of delay of feedback and have reported inconclusive results. Champion and McBride (1962) attacked the problem by employment of two levels of duration of delay feedback (2 and 5 sec) and two interpolated activities in a PA task with ANT method of presentation. Subjects either did nothing during the delay (N) or read words that were associates (A) of the stimulus presented immediately prior to the delay interval. The A groups were employed to examine an extension of Spence's (1956) hypothesis which assumes that a delay period allows the stimulus components to become attached to a variety of responses which compete with the response being learned, thus resulting in inferior performance. Analysis of speed (1/latency) of correct responses resulted in significant main effects of delay of feedback, practice, and activity. In addition, two-way interactions of practice with activity during the delay interval and practice with delay reached significant levels. As expected, delay of feedback and the reading of competitive response during delay decreased the speed of response and these effects became more pronounced as practice increased. The interaction of delay and activity was not significant indicating that the 5 second delay had adverse effects for both N and A groups.

Sampson (1971) extended the Champion and McBride design by including a third interpolated activity group in which the subjects read words dissimilar to the stimulus members and by reversing the delay duration after 14 practice trials for half of the subjects. In all other respects the two studies were similar. The results in terms of analysis of speed measures before the reversal of delay duration were comparable to those obtained by Champion and McBride. Performance was superior with the no activity group compared to the other two activity groups, which did not differ from each other. Superior performance was also found with the 2 sec delay when compared with the 5 sec delay. Moreover, these differences increased as practice increased. More evidence in support of delay effects was obtained in the analysis of post-reversal data. The change from 2 sec to 5 sec produces a significant decrement in performance and the change from 5 sec to 2 sec produced significant improvement in performance for the two groups with activity. However, for the no activity groups a significant decrement of performance was observed only following the change from 2 to 5 sec, with no improvement following shortening of delay duration.

Although these studies present strong support for the inhibitory effects of delay of feedback on PA performance, other investigators have reported no change or even superior performance with delay. Lintz and Brackbill (1966) reported a series of four experiments involving delay of feedback in PA and VDL. In the first two experiments, which differed only in terms of the task performed, delay of monetary reward for correct response was incorporated as well as delay of feedback. In the last two experiments, agains differing only in terms of the two verbal tasks, only delay of feedback was varied. In all experiments two levels of delay of feedback or monetary reward were included. These were 0 and 10 sec. Analysis of the number of errors to the criterion of three successive correct responses per item revealed a significant effect of delay in only one of the four experiments. The delay effect was only evident with the PA task and that only for feedback dealy in one of the two experiments. The immediate feedback group made significantly fewer errors than the 10 sec delay group. No delay effect was found in the experiments in which the VD task was employed.

Three investigators have reported improvement of performance with delay of feedback. Jones and Bourne (1964) employed a 3 x 3 factorial design consisting of three levels of delay of feedback and three levels of interval between the presentation of stimulus-response pairs and the presenatation of the following stimulus terms (designated as interunit interval, IUI). The durations of intervals for both variables were 0, 3 and 6 sec. The mean number of errors to the criterion of one perfect trial at 0 IUI level were 80.0, 60.9 and 53.4 for 0, 3 and 6 sec delay of feedback respectively. Analysis of this measure confirmed the significant superiority of performance with longer delays of feedback. In addition, a significant effect of IUI was reported, again indicating a decrement in errors, with longer intervals.

Similar results were reported by Nodine and Nodine (1966). In addition to delay of feedback and IUI variables, Nodine and Nodine varied the rate of anticipation and feedback exposures. Three durations (0, 1 & 2 sec) of feedback delay and IUI and two rates (½:½ & 2:2 sec) of anticipation and feedback exposure were employed. Unlike Jones and Bourne, however, a test trials was interspersed between each of the four acquisition trials to measure performance. Their results, based on an analysis of the number of correct responses on six test trials, revealed all three main effects to be significant. Number of correct responses increased directly with the increase of each of the three temporal variables. In addition, the data indicated that when duration of feedback and anticipation exposures were ½:½ sec increase of IUI from 0 to 2 sec produced significantly larger increments in correct responses than increase of delay of feedback interval. The reverse was true when duration of exposures was 2:2.

Jones (1968) also examined the effect of delay of feedback and IUI with the additional variables of modes of feedback presentation and types of interpolated activity during feedback delay. Feedback presentation consisted of either stimulus and response pairs or response terms presented alone, and the activity variable included no activity or counting numbers backward. The duration of interval for both delay of feedback and IUI were 0 and 6 sec. The effects of delay of feedback were greatly influenced by both the presentation mode and the nature of interpolated activity during delay. With immediate feedback performance was identical under both modes of presentation while the stimulus-response mode was considerably easier under conditions of delay of feedback. More importantly, the unfilled interval of delay of feedback with stimulus-response presentation facilitated performance, but no significant difference was obtained when response terms alone were presented. Counting during feedback delay decreased the level of performance to the point that no difference was found between counting delay and the no delay conditions.

The effect of type of feedback presentation on delay of feedback has also been studied in the VD task (Ward & Maisto, 1973). These investigators employed two durations of feedback delay (1 and 5 sec), two durations of feedback exposure (1 and 5 sec), and two types of feedback presentation (either the correct item was presented or its position was indicated). As in Jones' study, analysis of number of correct responses resulted in significant main effects of all three variables, but the trends were opposite to those reported by Jones. Superior performance was shown with the lower level of delay and with the longer duration of feedback exposure. Presentation of the correct response also increased the number of correct responses. The effect of delay of feedback, however, was only evident when the position of correct item was indicated during the feedback exposure. No difference between 1 and 5 sec delay was found when the correct item was presented.

The studies reviewed above indicate that the effects of delay of feedback are not consistent. Only two investigators (Champion & McBride, 1962; Sampson, 1971) reported inferior performance with delay of feedback in PAL. It should be noted that these investigators used speed as their dependent variable and the speed measure was based on latency of response only when subject has responded correctly. It cannot be determined if superiority of low delay groups also would be evident if more conventional measures such as errors or numbers of trials were employed.

A more puzzling result is the superiority of delay groups to no-delay groups reported by other investigators. Although the results can be perceived as a support for efficiency of learning with distributed practice, Underwood (1961) concluded that the literature prior to that time had shown little support for the superiority of distributed practice in verbal learning when the intertrial interval was manipulated as an independent variable. Moreover, Izawa (1971) reported no effects of distributed practice in PAL with the ANT procedure when the IUI was varied. One source of the superior performance with delay of feedback may be the rehearsal of items. The subject may rehearse the pair if his response during the anticipation period was correct or he could rehearse the pair presented to him during the previous feedback period by ignoring the presentation of the immediate stimulus when he is not able to retrieve the correct response. Some support for such a notion can be cited from results reported by Jones (1968). Jones obtained superior performance for delayed groups when no interpolated activity was employed but no facilitation was observed when subjects counted numbers backward during the delay period. Apparently the incompatible interpolated activity prevented rehearsal of items.

It can be concluded, therefore, that delay of feedback cannot explain the differences in performance resulting from MP for several reasons. First, the result of experiments indicating inferior performance with delay of feedback are contrary to the general finding that the ST method produces either superior (e.g., Battig & Brackett, 1961, 1963) or equal performance when compared to ANT method. Superiority of the ANT method over the ST method is rarely reported (e.g., Elmes & Lovelace, 1967) and can be regarded as the exception rather than the rule. Secondly, the superiority of ST cannot be explained by the results of studies reporting better performance with delay of feedback since the effect has been evident only when no interpolated activity is employed. In the ST procedure the feedback delay interval is filled with presentation of other stimuli or stimulus-response terms, prompting the subject to engage in storage or retrieval activities. The feedback delay period in the ST method, therfore, more closely resembles the conditions in which interpolated activity was employed. It should be remembered that superior performance was not observed in such conditions. Other theoretical views have been forwarded to account for the reported differences between the ANT and ST methods of presentation.

Theoretical Views

Three major theoretical views have been proposed to explain the resultant performance differences with the ST and ANT procedures. In an attempt to explain the superiority of ST procedure in their early study, Battig and Brackett (1961) assumed the complete temporal separation of the presentations of stimulus-response pairs from the presentation of stimulus terms alone to be the major source of facilitation. They felt that each ST trial "consists of two separate presentations of the entire list of S- terms, which might be expected to enhance the identification of the set of material to be learned, as well as to facilitate the differentiation of the items from each other and from material external to the list" (p. 63). It is assumed that this temporal separation in effect facilitates the division of the two behavioral processes of producing a previously learned correct response and learning of new stimulus-response associations. In the ANT method both these processes occur concurrently, leading to mutual interference and a consequent reduction of performance.

Battig and Brackett (1963) reported further support for their hypothesis. In an experiment consisting of three levels of intralist similarity, two MP, and two levels of percentage of occurrence of response members (%ORM), they found significantly superior performance with the ST method under 100% ORM, but inferior performance of ST group under 50% ORM, though not significant. They argued that when the separation of stimulus-response terms and stimulus-terms presented alone is not clearly distinguished, as is the case under 50% ORM, the advantage of ST will diminish. The inferior performance of subjects in the ST conditions to those in the ANT under 50% ORM was due to performance in only one group of subjects and was attributed to possible differences in samples of the population. In all other 50% ORM groups no difference was found between the two methods.

Battig and Brackett (1961, 1963) do not identify the mechanisms involved in the interference of recall and learning processes in the ANT method. Nor are they specific in terms of the effects caused by the temporal separation of learning and performance. In their earlier report they dismissed an extra 30 sec intertrial interval with the ST method as a possible variable affecting the outcome of their results. In a later report, however, Battig (1973) cites the intertrial interval as an important factor to account for the discrepancy between these experiments and those in which superiority of ST was not observed (e.g., Lockhead, 1962). This change of position was caused by Izawa's (1971) demonstrations of marked facilitation of paired associate ST performance produced by increases in either intra- or intertrial intervals. Izawa made no direct comparisons between the ST and ANT methods, nonetheless, her data clearly indicates the ST method to become increasingly superior to the ANT procedure with increases of either intra- or intertrial intervals. Battig (1973) argued that those studies finding superior performance with ST (e.g., Battig & Brackett, 1961, 1963; Cofer, Diamond, Olsen, Stein, & Walker, 1967, Experiment I) generally used relatively long intertrial intervals, while experiments finding no difference between the methods (e.g., Lockhead, 1962; Cofer et al, 1967, Experiments III & V) employed shorter intervals. It thus appears that Battig has shifted his emphasis on an "interference" hypothesis to a temporal one stressing the elapse of time between study and test trials as a strong facotr in facilitation of learning with the ST procedure.

Kanak and his associates (Kanak & Neuner, 1970; Kanak, Cole & Eckert, 1972; Duffy & Kanak, in press) have extended Battig and Brackett's original theory to include the confidence threshold of subjects as an important factor contributing to differences obtained with the two methods. The assumption is made that in the ANT procedure the subject withholds responding with an "available" item until he is relatively more confident of the correctness of his response. The setting of a higher confidence threshold is hypothesized to be a result of competition between the cognitive processing operation of storage and retrieval stemming from the rapid alternation between storage and retrieval in the ANT method. These assumptions were initially derived from the results of an experiment reported by Kanak and Neuner (1970). These investigators employed three variations of the ST procedure, depending on the item(s) that subjects were instructed to recall during the test trials. The subjects were either instructed to respond with the response terms (B) when the stimulus terms (A) were presented, ST-B; to respond with the "stimulus" terms when the "response" terms were presented, ST-A, or to respond with stimulus-response pairs (A-B) when a blank slide was presented to them during the test trials, ST-AB. In comparing the ANT method with the ST-A and/or ST-B procedures, it was found that the former resulted in a significantly higher rate of omission errors even though it did not differ from the latter groups in response learning stage measures. The results thus suggest that although the responses were equally available to the subjects in the ANT condition, the greater degree of withholding of responses may be correlated with a more stringent confidence threshold.

Further support for this hypothesis was reported by Duffy and Kanak (in press). Duffy and Kanak used two groups of subjects differing in their ego-strength as determined by the Sixteen Personality Factor Questionnaire (Cattell & Eber, 1962). One half of the subjects in each ego-strength group (i.e., high or low) learned a PA list by the ANT and the other half by the ST method. Of particular interest was the performance of subjects as measured by the number of trials until a response is first given (FG). Ekstrand (1966) suggested that FG measures the subject's readiness or willingness to emit a response and proposed that FG be termed a measure of the "confidence threshold." Duffy and Kanak predicted that high ego-strength subjects, who are assumed to be more capable of inhibiting overt response tendencies, should produce larger FG scores, reflecting a higher confidence threshold, under the ANT method. Low ego-strength subjects are less capable of inhibiting responding and, on the other hand, were predicted to show either no difference in FG scores between the two methods or produce smaller FG scores in the ANT method. The results of their experiment confirmed the prediction. High ego-strength subjects had significantly higher FG scores under the ANT procedure than under the ST method, while the low ego-strength subjects had smaller FG scores, though not significantly so, under the ANT procedure.

Although this hypothesis is similar to that advanced originally by Battig and his associates in that it allots a significant role to the competitive and interfering nature of the ANT procedure, it differs from the more recent temporal hypothesis of Battig (1973) in its lack of emphasis on the role of the intertrial interval and, more importantly, by its introduction of the confidence threshold concept. The confidence threshold concept stresses performance, as opposed to learning, processes in the two methods of presentation. The introduction of the confidence threshold concept has the additional advantage of enlarging the predictive power of the theory by assuming that all variables that may be instrumental in affecting confidence thresholds (e.g., intra- and interlist similarity, rate of presenatation) would also effect performance in various MP.

The most recent theoretical view is the "retention interval hypothesis" proposed by Izawa (1972). The retention interval hypothesis is concerned with the interval between the presentation of stimulus-response pair (reinforcement, R) and the presentation of stimulus alone (test, T). Since items are presented as either a R or a T on each trial, Izawa distinguishes four exhaustive parameters: R-R (inter-R-trial), T-T (inter-T-trial), T-R (feedback), and R-T (retention) intervals. The length of each of the intervals can be determined by the number of its intervening events of Rs and Ts in an _n_-pair list, assuming the rates of presentation are held constant under both the ANT and ST methods. Table 1 presents the means and ranges for length of each of the intervals under each MP as derived by Izawa (1972).

The R-R and T-T intervals, though having different ranges in the two methods, are identical in their mean under both methods and cannot be the major factors responsible for performance differences. The R-T interval (feedback delay) is always zero in the ANT method but has a mean of $\underline{n}$-1 under the ST procedure. If feedback delay was detrimental, the ST method should be inferior. As was already pointed out, this has not been borne out empirically.

Izawa sees a critical difference between the methods in the retention (R-T) intervals. Under ANT the shorest R-T interval occurs when a given item is presented last on trial $\underline{n}$ and first on trial $\underline{n}$+1, having 0 intervening events; and the longest one, when an item is first on trial $\underline{n}$ and last on trial $\underline{n}$+1 has 4n-4 intervening events. The other cases vary somewhere between the two boundaries with a mean of 2n-2. In contrast, under ST the range varies from 0 to 2n-2 with the mean of $\underline{n}$-1 events. The mean absolute R-T intervals for ANT and ST methods are expressed, respectively, as 2($\underline{n}$-1)d + (intertrial interval) and ($\underline{n}$-1)d + (intertrial interval); where $\underline{d}$ represented the rate of exposure and is assumed in both methods. Izawa suggests that the longer R-T interval filled with interference producing events constitutes a critical disadvantage for the ANT method by producing more short-term memory losses. The retention interval hypothesis thus states: "Other things equal, the RT [ST] method with the shorter mean retention interval with fewer intervening events should produce relatively better performance in acquisition than the anticipation method with longer interval having more intervening events" (Izawa, 1972; p. 18, original italics).

The distribution curves of $\underline{n}$ R-T intervals as a function of intervening events under a random item presentation are shown in figure 1. The closer an item is to the zero point of intervening events the less the short-term forgetting and the better the performance. The relative advantage of the ST method is obvious. More important is the large overlapping area under both curves, providing a potential explanation for the results of these studies in which no difference between the two methods were obtained.

A test of the retention interval hypothesis is provided by an experiment conducted by Izawa (1972) in which the R-T interval was manipulated under both ST and ANT procedures. A list of 20 CVC-two digit number pairs was divided into two sublists (X and Y) of 10 pairs. In one condition under the ST method, the R-T interval for items in the X sublist was shortened by placing the X items at the last half of each study trial and the first half of each test trial. The items in the sublist Y obviously had a relatively longer R-T interval. A similar procedure was employed under the ANT method so that in some trials the X items, and in others the Y items, had short R-T intervals. Two control groups, one for each of the two methods, were constructed in which the relative length of the R-T interval for both sublists was equal. It was predicted that in the experimental condition under the ST method the X sublist would result in less incorrect responses than the Y sublist. Similar predictions were made for the short and long R-T interval items in the ANT procedure. As expected, significantly more incorrect responses were produced for the Y sublist than the X sublist in the experimental condition of ST procedure, while no difference among the two sublists were found in the control condition. When the Y and X sublists were combined the experimertal and control conditions under the ST method resulted in equal performance. Similarly the performance was found to be a function of the R-T interval in the ANT method. Furthermore, the performance under the combined ST conditions was significantly better than under the combined ANT conditions. These results provide strong support for the retention interval hypothesis.

It is important to note that Izawa stresses short-term memory as a functional variable contributing to the differential effects of the two methods. It is predicted, therefore, that the recall of items after a relatively long retention interval would be equal under ST and ANT method. This prediction was confirmed in a second experiment (Izawa, 1972; Experiment II) in which an additional trial (a test trial in ST condition) was presented after 15 minutes of unrelated interpolated activity. The superiority of the ST method evident during acquisition diminished in the recall measures.

Although Izawa allots some importance to the interfering effects, the interference is not attributed to the alternating storage and retrieval processes as suggested by the Battig and Kanak research groups, but rather to the number of intervening events regardless of the processes involved during these events. Izawa's hypothesis also lacks the emphasis on the temporal separation (i.e., intertrial interval) and stresses the function of intervening <u>events</u> (Izawa's personal communication to Battig, see Battig, 1973).

Izawa does not specify the mechanisms underlying short-term memory losses. Tulving and Arbuckle (1963) identified two sources of interference in short-term recall. One source, referred to as output interference, is produced by retrieval of new material from memory; and the other, input interference, is caused by the input of additional items in memory. Both types of interferences could be operative during the retention interval. Inclusion of these concepts in the retention interval hypothesis could enhance the theory in terms of both its explanatory and predictive powers.

Battig (1973) has attacked the retention interval hypothesis on a number of positions. Battigs argues that Izawa's contention that the overlapping areas in the distribution curves of the R-T interval under the two methods of presentation can account for inconsistent results is falacious. Individual ANT pairs which overlap the length of the retention interval with individual ST pairs have to be compensated for by other ANT pairs with longer retention intervals than any ST pairs. Since Izawa assumes the retention interval length to be the only factor contributing to the differential performance between the methods and since the mean retention interval depends only on the number of items in the list, there seems to be no way that the retention interval hypothesis can account for the discrepancies in the literature. Battig alos argues against Izawa's disregard for the effect of the intertrial interval. Using Izawa's own formulations, 2(n-1)d + (intertrial interval) for ANT, and (n-1)d + (intertrial interval) for ST, Battig demonstrated that a large intertrial interval would increase the absolute retention interval differences in the two methods and suggests the different intertrial intervals employed in different experiments as the main factor contributing to the inconsistent results in previous studies.

Although the PRO method of presentation is treated as a variation of St procedure and is reviewed as such throughout the present paper, it is of interest to examine some theoretical explanations for the superiority of the PRO method over ANT advanced by Cook and Kendler (1956) and Peterson and Brewer (1963). Cook and Kendler conducted an experiment consisting of two methods of presentation (ANT & PRO) and two levels of training. Subjects in the high level of training received four pre-training trials during which the stimulus and response terms were presented but subjects were not tested for learning. Subjects in the low training condition had no pre-training trials. In both conditions subjects received three experimental practice trials. In the PRO procedure, the stimulus term was presented first followed by the presentation of response term, after which subject wrote the response term. The anticipation method differed from the traditional method in that the response term was presented alone during the feedback period. The dependent variable was the number of correct responses on a test trial presented following the termination of practice trials. The PRO procedure was found to produce significantly more correct responses than ANT regardless of the level of training. To account for this finding, Cook and Kendler proposed a two-stage hypothesis of PAL. They argued that in order to establish an association between a stimulus and response term, subjects first produce an implicit cue-producing response to each of the two items. These implicit responses are assumed to be pre-experimentally learned. After these implicit responses are established, an association between the implicit response to the stimulus term and the implicit response to the response term is formed to complete the learning of the nominal pair. The first stage of PA learning thus consists of formation of an implicit response to the two items and the second stage involves the formation of an association between the two implicit responses. Cook and Kendler argued that the PRO method results in superior performance because it facilitates the accomplishment of the second stage of PAL. This facilitative effect is attributed, at least in part, to the short delay between the presentation of stimulus and response term, omission of overt response, or both, in the PRO method. Cook and Kendler, thus, assume that overt practice or "performance" interferes with associative formation.

Peterson and Brewer (1963) also attribute superiority of the PRO method partly to omission of overt responses. However, their formulation differs from that of Cook and Kendler in that they allot an important role to the interfering effects of incorrect responses in the ANT procedure as the major factor in hindering learning by that method. In addition, Peterson and Brewer have proposed the number of stimulus and response pairings per trial as another factor affecting learning under the two methods. In the PRO procedure, two such "pairings" occur in each trial (one due to presentation of items and one due to the pronunciation, as subject always responds correctly), while in the ANT method the "pairing" is limited to one per trial (just presentational). Accordingly Peterson and Brewer (1963, Experiment II) employed three variations of the PRO procedure to examine their hypothesis. In the first condition only one pairing of stimulus and response terms was presented in each trial; in the second condition two pairings and in the third condition interference was introduced by pairing of each stimulus with two different responses and instructing the subject to respond with the last response paired with a given stimulus. Learning was found to be less efficient in the interference condition than the other two conditions which did not differ from each other. Peterson and Brewer concluded that the inferiority of the ANT method is caused by interfering incorrect responses. The comparison of interference produced by a double function list in the PRO method, however, to the interference caused by an incorrect response in the more commonly used single function list in the ANT method is hardly justifiable. What the results of Peterson and Brewer show is merely that interference hinders learning, a finding which had been established long before. The conclusion that the inferiority of the ANT method is due to interference of incorrect responses is well beyond the scope of their experimental design and data.

The studies reviewed in the present paper are evaluated in terms of the evidences they provide in support of, or against, the theoretical positions described. It should be noted, however, that many of the investigations reviewed were not designed to test these theories and their evaluation in such terms is admittedly ad hoc. As will become clear later in this paper, after a more thorough review of the literature, none of the three theoretical formulations discussed above can explain all the empirical results reported.

Task Variables

The main verbal tasks used in the examination of the effects of method of presentation are PAL, SL and VDL. Of these, by far the most common task employed is paired associate learning followed by verbal discrimination and serial learning in that order. Unfortunately, no single investigation of the effect of methods across different tasks with common materials is available. To avoid repetition, this section is devoted solely to those studies in which no secondary variable is manipulated. Detailed discussion of the effects of secondary variables is reserved for a later part of this paper.

Paired Associate Learning

Soon after Battig and Brackett (1961) reported superiority of the ST procedure to the ANT method in PAL, Lockhead (1962) presented results in contrast to those of Battig and Brackett. Lockhead expanded Battig and Brackett's design to include a third method of presentation and a separate variable of partial knowledge of results. In addition to the ST and ANT methods, a variation of anticipation (random, RAN) procedure was used in which the stimulus-response pairs presented during the feedback period did not correspond with the stimulus terms immediately preceding them but were randomly placed in that position. The RAN condition, therefore, incorporated both the alternation of storage and retrieval processes, a characteristic of the ANT method, and the delay of feedback inherent in the nature of the ST method. In addition, half of the subjects in each presentation condition were given oral knowledge of result ("right" or "wrong") immediately after their response and the other half had no feedback. Based on the results of animal studies, Lockhead predicted superiority of ANT to the other presentational conditions and a greater effect of knowledge of results in the ST and RAN conditions than in the ANT. The mean number of trials to the criterion of one errorless trial for ANT, ST and RAN conditions were respectively 24.2, 23.6, 28.3 for the groups with knowledge of results and 25.3, 27.7 and 28.0 for groups without knowledge of results. As the means show, neither of the two main effects nor their interaction produced significant differences.

Although there were numerous methodological differences between the two experiments (e.g., rate of presentation, material, list length, to cite a few); Lockhead attributed the discrepancy in the results only to the differences in the material used. Battig and Brackett employed nonsense shapes as stimuli while Lockhead employed CVCs. Lockhead referred to results presented by Gagne (1950), who found that associations are learned faster when members of the same nonverbal stimulus class are presented sequentially, in support of his argument and suggested that the ST method can facilitate stimulus discrimination when such materials are used. Whether or not material selection has a significant effect on the outcome of method of presentation effects will be examined in detail later in this paper. The importance of Lockhead's results, however, lies in the finding that the superiority of ST over ANT method is not universal.

Results similar to that of Lockhead were reported by Cofer et al (1967; Experiment VI). In addition to the three methods employed by Lockhead, Cofer et al included a fourth called "doublet," similar to the RAN method in all respects except that in the doublet method a set of two stimulus terms and stimulus-response pairs alternated. Again no significant difference in learning rate was found among the four methods. Cofer et al found these results rather puzzling since in an earlier experiment (Cofer et al, 1967; Experiment I) using the same material but a shorter list and only the ST and ANT methods, significant superiority of ST method was obtained. Moreover, in another experiment of this series (Experiment IV), in which a very short list was employed, the ANT method produced superior performance in comparison to the ST method. These findings prompted the authors to conclude that "the advantage of faster learning under recall [ST] than under anticipation is inconsistent" (p. 556), and that the results obtained by Battig and Brackett (1961) are not robust.

Although their first and sixth experiments differed in that the ST condition in Experiment I had an extra 30 sec study-test trial interval, Cofer et al disregarded this variance as a source of explanation of the discrepancy in results on the base of the general lack of any effect of intertrial interval on PA learning (Underwood, 1961). As mentioned earlier, Battig (1973) has argued that the difference in intertrial interval is indeed the main source of discrepancy, not only in these two experiments, but perhaps throughout the reported literature.

Discrepant results between methods of presentation have been the rule rather than the exception when the ST and ANT procedures of PAL are compared. Of the 37 reports of such comparisons available to this reviewer, 24 have reported significant superiority of the ST method either as a main effect or through interaction with other variables, while 13 have been unable to find any differences. Only a few investigators (Cofer et al, 1967; Elmes and Lovelace, 1967, Goss and Nodine, 1965) have ever reported superiority of ANT over the ST method.

The inconsistent effect of MP is reflected in the studies reviewed in the present section as well as throughout the paper. For example, when Rothkopf and Coke (1963) compared a variant of ST and two ANT methods in a within-subject design on a mixed list of connected discourse, no difference in learning due to method of presentation was found. Battig and Wu (1965) attributed the lack of an effect in Rothkopf and Coke's study to their use of an unpaced rate of presentation and suggested that with such a procedure the separation of learning and performance may become meaningless. Accordingly, Battig and Wu employed a paced PA task in which the two methods of presentation were mixed in a single list, a procedure similar to that used by Rothkopf and Coke. A 12 pair list was divided into six ST and six ANT pairs. For one group, separate pairs (P), stimulus terms of the ST pairs were intermixed with the ANT stimulus terms and stimulus-response pairs, all of which were followed by the six ST stimulus-response pairs. In the other group, separate test (T) lists each started with the presentation of six ST stimulus-response pairs intermixed with ANT pairs and stimulus terms and proceeded to the presentation of ST stimulus terms. In both groups the presentation of stimulus terms of ANT pairs were immediatley followed by their corresponding stimulus-response pairs. It should be noted that the P sequence involved temporal separation of ST stimulus-response pairs but not ST stimulus terms, while the reverse was accomplished by the T sequence. Further temporal separation of ST stimulus-response pair or stimulus terms was provided under the interval

(INT) conditions with an unfilled 15 sec blank interval both preceding and following these presentations, whereas in the continuous (CON) conditions an uninterupted 4 sec rate was employed both within and between each of the 12 successive trials.

In contrast to the results of Rothkopf and Coke, overall superiority of ST pairs to ANT pairs was highly significant. Moreover, a marked overall superiority of the P conditions over the T conditions and a marginal facilitative effect of INT condition was obtained. These results indicate that not only separation of learning and performance is an important factor in learning efficiency, but also the separation of the stimulus-response pairs produces more effective learning than separation of stimulus terms, or using another terminology, storage mechanisms are more sensitive to interference than are the retrieval mechancisms. Such results obviously provide strong evidence in support of the hypotheses proposed by the Battig or Kanak research groups, but fail to identify the source of discrepancy in the results of the two experiments. Since Battig and Wu did not include pacing (i.e., pace versus unpaced) as a variable, the study fails in its purpose to demonstrate the cause of the contradictory results in view of the numerous other differences in the methodology of the two experiments (e.g., material).

Support for the supremacy of the ST method is also reported by Moursund and Chape (1966). Their study consisted of three MP. One was the traditional ANT procedure, the second was a variant of ANT in which subjects had a list of all the ten responses available througout the task. In both these methods a 10-pair PA list was presented at a 3:3 sec rate. The third method differed slightly from the traditional ST procedure in that subjects were given a list of all the ten pairs for a period of 30 sec which constituted a study trial. No difference was found between the two ANT methods but significant superiority of ST to the other two methods was confirmed. In addition, an interaction of MP with trials indicated that the difference between the ST and ANT methods increased with the number of practice trials. These findings confirm the results reported by Kanak and Neuner (!970) in which the difference among the ST and ANT methods was attributed to omissions of responses during the associative learning stage. The equal availability of response terms (no difference in response learning stage) did not produce equal associative learning, as the ANT method continued to produce a higher rate of omissions, suggesting that a higher confidence threshold is developed by this procedure.

Although the superiority of PRO over ANT method in PAL has been shown to be more consistent, discrepancies between the results still exists. Of the 18 experiments available, 14 have confirmed the results of Cook and Kendler (1956), while 4 have found no difference between the two procedures. Furthermore, the superiority of PRO over ANT seems to be limited to the early learning trials, specifically when multiple choice PAL is employed as a task (Hawker, 1964, 1965a, 1965b).

Cook replicated the Cook and Kendler (1956) study employing a greater number of practice and test trials. It will be remembered that only three practice and one test trials were presented in the earlier study. Under such a procedure the highest degree of correct responding achieved was about 58%. Cook (1958) questioned the conclusion of superiority of PRO over the ANT method with such a low level of learning and argued that had the training been continued further the ANT method might have proved superior in the later stages of learning. Accordingly, 36 practice and

12 test trials were administered to subjects. In all other respects the two experiments were identical. Significant superiority of PRO method was confirmed in the first nine but no difference was present in the last three testing trials. The lack of difference in the late test trials solely reflects perfect performance of subjects in both conditions during these trials.

Peterson and Brewer (1963, Experiment I) also reported superior PRO performance in a within-subject design. Their procedures differed with those employed by Cook and his associates in that Peterson and Brewer employed alternating test and practice trials (compared to 3 practice and one test rials) in PRO method. Practice was carried out to 15 trials. The PRO method produced superior learning in the first five trials but, in agreement with Cook (1958), no difference in performance was observed in the later trials.

More evidence for superiority of PRO over ANT is reported by Kopstein and Roshal (1955). In their version of the PRO procedure stimulus and response terms were presented simultaneously for a period of 3.5 sec each, while ANT method consisted of traditional anticipation and feedback period with a total time span equal to that in the PRO method. Their result indicated that the PRO procedure was significantly superior in the early trials but the advantage of PRO became distinctly less apparent when higher levels of learning were reached.

These results prompted Sidowski, Kopstein and Shillestad (1961) to suggest that more efficient learning may take place when the subject merely observes the presentation of stimulus and response terms than when he responds. Such an argument is similar, but more general in its scope, to Peterson and Brewer's hypothesis that incorrect responses interfere with learning. To test their hypothesis, Sidowski et al included six different presentational procedures. These conditions differed from each other in practice trials only, having identical test trials. In their PRO method the stimulus was presented for one sec followed by the presentation of the response term for 4 sec which preceeded a 6 sec interval during which the subject wrote down the response term. The PRO-SR procedure was identical to the PRO method with the exception that both stimulus and response terms were presented simultaneously, instead of the response term alone, maintaining the same rate of presentation. The ANT method consisted of the presentation of stimulus term for one sec, a 6 sec interval during which subject anticipated and wrote the response term followed by a 4 sec presentation of the response term alone. The ANT-SR procedure was identical to the ANT condition with the exception of simultaneous presentation of stimulus-response terms in place of the response term alone. In the ANT-PRO procedure, standard ANT procedures were used, however, the subject was required to write down the response term during the feedback period as well as in the anticipation period. Finally, in the Simultaneous observation (SIM) condition stimulus and response terms were presented for 5 secs with a 6 sec interval during which subject made no overt or written response. All subjects were given three test trials, on after the sixth study trial and two after the tenth. In the first two test trials stimulus terms were presented for one sec each and subject was provided 4 sec to write the response terms. In the third test trial response terms were read to the subject and he was instructed to provide the stimulus term. Table 2 presents the mean number of incorrect responses (out of a possible 15) on each of the three test trials. Analysis of this measure produced the following results. On the first test trial the PRO-SR condition was superior to all other conditions except the SIM which itself was superior to all groups but PRO-SR and PRO. On the second test trial again PRO-SR procedure resulted in significantly fewer errors than all other groups but SIM and PRO; SIM was superior to all the ANT conditions and PRO was superior to ANT. In the third test trial, reflecting a backward recall test, the only significant difference was due to better performance of subjects in the SIM condition than those in the ANT conditions. These results only partially support the investigators' hypothesis that observation of stimulus and response terms alone results in most efficient learning. Fewer errors were produced in the PRO-SR condition than in the SIM condition, though the difference was not significant. The results can be explained best in terms of the rehearsal time provided. In conditions PRO, PRO-SR and SIME subjects have a period of 10 (or 11 in SIM) sec to rehearse each pair, while in the three remaining conditions only 4 sec is allotted to rehearsal. The results also show that simultaneous presentation of stimulus-response terms results in better learning. It is not clear whether simultaneous presentation aids the associative learning or stimulus discrimination component of PA learning. It should be noticed that the SIM procedure is similar to the ST procedure in that during training trials stimulus-response terms are presented together and superiority of this condition over the PRO condition can be conceived of as comparable to more efficient learning with the ST procedure.

Comparison of PRO, ANT and SIM procedures is also reported by Reynolds (1967; Experiment I). In addition to MP, Reynolds also examined the effects of number of test trials on PAL. In one condition, test (T) trials were administered after the third and the sixth trials. In the other conditions, no test (NT) trial was presented before termination of training trials. Subjects in both conditions received five test trials following termination of practice trials. Two dependent variables were used in the analysis of the data. These were the total number of correct responses over (a) 5 testing trials and (b) on the first test trial after the termination of training. No significant effect was obtained with the analysis of the former measure. However, the main effect of number of test trials was significant and its interaction with MP approached significance when the results of the first testing trial was considered. Subjects in the T groups performed better than those in the NT groups. The interaction of MP and number of test trials indicated the lower performance level of subjects in the SIM-NT condition. This finding is clearly in contradiction of that reported by Sidowski et al. in that the SIM condition was not superior to PRO or ANT conditions and indeed showed inferior performance in one testing condition. Moreover, the superiority of PRO over the ANT procedure typically found in other studies was not observed.

The lack of an MP effect in a comparison of PRO and ANT procedure of PA learning is also reported by Hawker (1967). Hawker employed a 2^3 factorial design with two MP (PRO & ANT), two forms of PA (traditional and multiple choice) presentation, two locus of forms of PA presentation (during test or training trials). The multiple choice presentation consisted of presenting a stimulus with four responses, all of which were included in the list but only one of which was paired consistently with a given stimulus term. Multiple choice presentation produced better performance than the traditional PA task; however, no differences in the MP (PRO vs ANT) were obtained.

As in the investigations of the ST procedure, the numerous methodolgoical differences does not allow direct comparison of these studies and makes it impossible to pinpoint the major potential sources of discrepant results.

Verbal Discrimination and Serial Learning

Comparisons of MP in tasks other than PAL are scarce. This review could find only six reports of investigation of ANT and ST and one report of examination of ANT and PRO in VDL and a total of four investigations (three concerning ST and one PRO) in SL. In general the results of these investigations of VDL and SL are more consistent than those reported with PA learning. This is specifically the case with the ANT and ST methods of presentation comparison. The superiority of ST and the PRO procedure over the ANT method is demonstrated in every study (i.e., Battig & Switalski, 1966; Fulkerson & Johnson, 1971, in VDL; Battig & Lawrence, 1967, in SL) with the exception of one for each task (Posnansky, 1972, in SL; Rowe & Paivio, 1972, in VDL). Better verbal discrimination performance with ST or PRO has been found in a variety of material, including connected discourse (Silberman, Melaragno & Coulson, 1961); different rates of presentation and across a number of other variables.

The reason for the more consistent results with VD learning in contrast to the inconsistent results obtained with PA learning cannot be easily explained. One would expect that separation of storage and retrieval processes would be more beneficial in PA learning since recall mechanisms are generally considered more complex than those of recognition. Furthermore, the threshold for recognition responses is typically assumed to be lower than that of recall responses. If superiority of ST over the ANT method is caused by relative increase of the confidence thresholds with the ANT procedure, as Kanak and Neuner stated, the effect of MP should be more evident in PA than in VD learning. One factor in the two tasks that may be responsible is the higher susceptability of recognition responses to greater interference from the number of intervening events. Since in the ST method the number of intervening events between feedback and anticipation of any item is, on the average, one half as much as it is under the ANT procedure, as Izawa (1972) pointed out, it is conceivable that the facilitative effects of ST may be greater in VD than in PA learning for this reason. It should be noted, however, that the effects of differential numbers of intervening events on VD learning has not been examined and the generalization of the effects of this variable on single-item recognition learning (Martin & Melton, 1970) to VDL may prove to be premature.

The superiority of ST over the ANT procedure in SL is more easily explainable. In the ST procedure separation of study and test trials provided subjects with twice as much time to retrieve or store the material as in traditional SL by the ANT method. This additional time undoubtedly is instrumental in producing better performance under the ST method. In none of the investigations has this time factor been controlled, as is essential for any meaningful comparison of the two methods in serial learning.

Method of presentation has been reported to have other effects on SL. Battig (1969) reported a marked reduction of classical serial position curve effect when the ST procedure was employed and suggested that the well-known and carefully studied serial position effect may be just an artifact of the ANT method of presentation. In addition, Posnansky (1972) reported that subjects under the ANT method of SL were more prone to use a multiple cue of three-prior-items while subject under ST made more effective use of a single prior-item cue. Such results may indicate the more complex nature of learning under ANT, perhaps as a result of interference between retrieval and storage processes.

The only valid conclusions that can be made in regard to the effects of MP on different verbal tasks is that, for whatever reason, the superiority of ST over the ANT method in PA learning is not universally found, but is more often than not, and the superiority is more consistent with VDL and SL tasks. Other variables may be instrumentally effective in affecting the outcome of the results. The next section of the present paper is devoted to review of the effects of these secondary variables.

SECONDARY VARIABLES

Material

The effects of the nature of material used in list construction on MP has been investigated more extensively than any other variable. The reason for such high interest in this variable is not clear, as only one of the major theoretical views, that advanced by Kanak and his associates, predicts differential performances in the three MP with different materials. However, the bulk of the research conducted predates the theoretical controversy and correlates with the emphasis on attributes of materials as they differentially affect learning as predicted by stage analysis concepts (Underwood & Schulz, 1960). As noted earlier, Kanak and Neuner (1970) attributed the advantage of ST procedure to the omission of responses in the associative learning stage of PA learning. It might be expected then that response term meaningfulness should not differentially affect learning under different MP. However, since the use of low meaningful responses requires a greater proportion of learning time to be invested in response learning stage, the facilitative effect of ST method may be over-shadowed when such response terms are employed, thus attenuating the difference in the two methods. It should be noted, nevertheless, that Kanak and Neuner employed highly frequent words as response terms. The use of such material may be responsible for the lack of difference in the response learning stage under the two methods. Low meaningful, unintegrated response terms may have an even greater effect on confidence thresholds under the ANT method and extend the differences between the ANT and ST methods to the response learning stage. Thus, the inferiority of the ANT method could be evidenced in both the response and associative learning stages when low meaningful responses are used, resulting in accentuation of MP differences. Since stimulus meaningfulness has been shown to have little effect on PAL, its effect on MP is expected to be minimal.

The effect of varying levels of stimulus and/or response similarity, may also interact with methods. Underwood, Runquist and Schulz (1959) demonstrated that response similarity has a facilitative effect on the response learning stage and an adverse effect on the associative learning stage of PAL. Likewise the hindering effect of stimulus similarity has been generally attributed to its interference with associative learning processes (see Kausler, 1974). The focus of effects is thus shifted from response learning with meaningfulness to the associative learning stage with similar stimulus and/or response terms. In addition, Ekstrand (1966) reported that the confidence threshold, as measured by FG, increases with intratask competition. Accepting the assumptions forwarded by Kanak and Neuner, one can predict that an increase in stimulus or response similarity would have a greater effect on ANT than on the ST method of presentation and results in greater superiority of ST over ANT. Moreover, manipulation of conceptual response similarity would be expected to facilitate the response learning stage under both methods, but the relative degree of facilitation under the ANT method may be somewhat reduced by the intrinsic source of competition within the method. Hence, only a subtle effect might be observed in the response learning stage, but a more potent difference between methods in the associative learning stage. On the other hand, manipulation of formal response similarity would increase the relative inferiority of the ANT method not only in associative learning stage, but also extend its effect to response learning stage.

The same line of reasoning can be presented for the effects of imagery on MP. Paivio (1969) explained the positive effects of imagery on PAL in terms of an encoding of the stimulus and response units of a pair into a compound image. That is to say imagery facilitates the associative connection of the stimulus and response terms. Moreover, the effect of stimulus imagery has been shown to be greater than that of response imagery (Paivio, 1965). Low stimulus imagery, therefore, is expected to elevate the performance differences between ST and ANT procedures.

The expected effects of verbal attributes on MP in VDL is less clear. The frequency theory of verbal discrimination (Ekstrand, Wallace & Underwood, 1966) explains VDL in terms of the differential frequency build up between the right and the wrong items of a pair. The frequency theory predicts an increase in the learning rate with an increase in the frequency of occurrence of right items and a decrease in the frequency of occurrence of the wrong items. The theory, with the incorporation of the Weber's Law analogy, also predicts better performance when both items of a pair are low in frequency rather than when they are high in frequency. However, the frequency theory makes no specific predictions concerning the effects of meaningfulness and imagery on VDL. Although the empirical results of experiments varying meaningfulness on VD learning are inconclusive, the general effect of the imagery attribute has been that of facilitation (Eckert and Kanak, 1974).

Insofar as the PRO procedure is similar to the ST method and in view of the fact that Peterson and Brewer (1963), as well as Cook and Kendler, have attributed the inferiority of ANT method relative to the PRO method to associative interference, the same theoretical predictions regarding the effects of verbal attributes of the materials made for the ST procedure can be extended to the PRO method.

Meaningfulness. Investigations concerning meaningfulness, frequency of occurrence and pronounciability are all presented in this section because of the similarity of the effects of these variables on verbal learning tasks. Although an interaction of MP and meaningfulness in PAL is reported in a number of investigations (Elmes & Lovelace, 1967; Goss & Nodine, 1965; Wright, 1967), the effect is by no means free from inconsistencies. Battig, Brown and Nelson (1963; Experiment IV) found no such interaction when they varied stimulus and response meaningfulness facrorially. Low (L) meaningful items consisted of CVCs with midrange difficulty and high (H) meaningful items were CVC words with an associative value of 100% (Archer, 1960). In addition to the H-H, H-L, L-H and L-L groups, a mixed list group was included in which subjects learned a ten-pair list with half of the pair comparable to those used in the H-H condition and the other half similar to those employed in the L-L condition. Although the reliable advantage of ST over the ANT procedure and the expected main effects of meaningfulness were obtained, the interaction of the variables did not reach accepted levels of significance with any of the dependent variables employed.

Similar results were reported by Goss and Nodine (1965; Experiment

7). Their study was identical in design to that of Battig et al. (1963) with the exception of the lack of the mixed list group. Goss and Nodine used CVCs with a meaningfulness value of 1.43 or less (Noble, 1961) for the L meaningful items and trogram words with values of 4.01 or higher as the H items. Although performance under the ST procedure was consistently better than under the ANT method, the analysis of the data produced no significant MP effect. The only reliable effects obtained in term of trial to criterion were those of stimulus and response meaningfulness.

A replication of this study (Goss & Nodine, 1965; Experiment 11) produced similar results. However, an interaction of meaningfulness and MP was revealed when a mixed list design was employed (Goss & Nodine, 1965; Experiment 9). Using the same materials as in their other two experiments and a 12-pair list ocntaining three pairs from each of the four L and H conditions, Goss and Nodine not only found a significant superiority of ST but also an interaction of MP with meaningfulness indicating a greater superiority in performance of H-H over L-L pairs under ANT than under ST method. Such results are in accord with the prediction of the competition hypothesis of Kanak and Neuner. It should be noted, however, that the advantage of ST over ANT method was evident over all levels of stimulus and response meaningfulness, a result in contrast with the competition theory of MP which predicts that low meaningful materials produce a more potent MP effect.

The source of the discrepancy of results between mixed and unmixed list design is not clear. The discrepancy cannot be attributed to an artifact of experimental procedures since it has been obtained by other investigators. Wright (1967), for example, also found a differential effect for MP in a mixed list design. Stimulus meaningfulness was varied in this study with L meaningful stimuli having an average associative value of 10% and H meaningful stimuli 90% on Archer's scale. Both mixed and unmixed list designs with respect to MP were employed and for the unmixed groups meaningfulness was a within-subject variable. As in the Goss and Nodine study, substantially better performance under the ST method was obtained in the mixed list design, but no difference in MP was observed with the unmixed list design. However, in contrast with the earlier experiment, interaction of meaningfulness with MP was obtained only under the unmixed design. The interaction reflected the greater superiority of ST over ANT procedure under low stimulus meaningfulness.

Differential effects of meaningfulness on MP was also reported by Elmes and Lovelace (1967). These investigators only used H-H and L-L groups in an unmixed list design. The selection of items was very similar to that reported by Goss and Nodine. Elmes and Lovelace found the performance under the ANT method to be consistently better than that under the ST procedure with the higher level of meaningfulness. The difference, however, only approached a significant level. On the other hand, performance was better under ST than the ANT method with the lower level of meaningfulness, although this difference clearly was not significant. Elmes and Lovelace attributed the discrepancy of their results with those of Battig et al. (1963) to the wider range of meaningfulness employed in their study. However, this argument cannot explain the differences in the results of their experiment with those of Goss and Nodine's. Although the superiority of the ST method under the low level of meaningfulness is in agreement with the prediction of the competition theory of Kanak and Neuner, the significant superiority of ANT over ST under the H-H condition cannot be explained either by the interference or any other theoretical position.

The discrepant results of these experiments are especially puzzling in view of the similarity of methodological procedures among them. All experiments used lists of medium lengths (ranging from 8 to12 pairs), comparable rates of presentation (2:2 to 4:4 sec), and similar intertrial intervals (5 to 10 sec).

Of the three studies (Reynolds, 1964; Levine, 1965; Hawker, 1965a) investigating the comparative effect of meaningfulness on the PRO and ANT procedures in PA learning, none have found an interaction of the two variables. Reynolds reported two experiments in which response pronounciability was varied. The two experiments differed only with respect to list design. The first employed an unmixed and the second a mixed list design with regard to pronounciability. The H pronounciability items had a mean of 2.46 and the L items a mean of 5.56 on the Underwood and Schulz (1960) norms. Only the effect of pronounciability was found reliable in the first experiment. Neither the expected superiority , derived from Cook and Kendler or Peterson and Brewer's hypotheses, of the PRO method nor the interaction of the two variables was observed. In fact, the small difference obtained between the two methods was in favor of the ANT method. However, a reliable advantage of the PRO method was obtained when a mixed list was used. These results are similar to those reported by Goss and Nodine and Wright with the ST method and indicate that the assumed higher level of interference effects in a mixed list generalizes to PRO-ANT comparisons.

Levine (1965) varied stimulus and response meaningfulness factorially. For the L meaningful items CVCs with a mean associative value of 74% (Glaze, 1928) and for H meaningful items two digit numbers were used in an unmixed list design. Although the main effect of MP and response meaningfulness were significant, the effect of stimulus meaningfulness and interaction of variables were lacking. Performance was more efficient under PRO procedure and with H meaningful responses. Hawker (1965a) reported similar results with a multiple response choice PAL task. In addition to response meaningfulness, Hawker manipulated availability of the response term. In the L meaningful response list stimulus terms consisted of adjectives and response terms of nonsense figures. For the H meaningful response list the stimulus and response terms were interchanged. The alternative choices were limited to 8 responses within the list in the L response availability conditions and to 16 responses, including 8 extra-list items, in the H response availability conditions. Hawker reported a significant advantage of PRO over ANT method only in the early learning trials, and although the effects of response meaningfulness and availability were both highly reliable, their interactions with MP were not.

It seems apparent that the superiority of PRO is not dependent on the meaningfulness of stimulus and response terms. It is highly conceivable, however, that the advantage of a longer rehearsal time in the PRO procedure overshadows any possible differential effect of meaningfulness on the two MP.

The investigations of the effects of meaningfulness on MP in a VD task are scarce. Only one such investigation has been reported. Ingison and Ekstrand (1970) manipulated frequency and imagery of the items in a comparison of the ANT and ST methods of VDL. Four sets of 5 pairs each were used representing each of the four conditions resulting from factorail manipulation of imagery and frequency in a mixed 20-pair list. Low frequency items had a frequency of 15 per million or less and H frequency items were rated 45 per million or more on the Thorndike and Lorge scale. Imagery measures were drawn from Gorman's (1961) compiled lists. The effect of MP in favor of the ST procedure was found to be reliable, however, neither of the word attributes had any effect on the rate of VD learning.

In conclusion, the effect of meaningfulness on MP of verbal tasks can best be characterized as inconsistent. In the single study reported on VDL, manipulations of frequency produced no differential effect on learning efficiency under the two MP. In comparisons of the ST and ANT procedures on PAL the results, when the differential effect of meaningfulness is obtained, are in support of the competition theoretical position. The early conclusion of Battig and Brackett (1963) that the superiority of ST method is consistent over a wide range of stimulus and response meaningfulness is no longer valid in view of results reported by Elmes and Lovelace (1967) and seems to have been premature.

Interitem Similarity. The results of experiments investigating the effects of stimulus and response similarity are also inconclusive. Of the seven studies reported in PAL, four have demonstrated some interaction of interitem similarity and MP while three have found no interaction. In a study reviewed in an earlier section, Battig and Brackett (1963) covaried stimulus and response similarity in a VD to PA transfer task design. Three levels of similarity were employed. A 12-pairs CVC list was used for all groups. Subjects in the H similarity conditions learned a PA list consisting of stimulus and response terms constructed from four consonants and three vowels. In the list for medium similarity conditions the stimulus and response terms were composed of 8 consonants and three vowels while the list for the low similarity conditions contained all consonants and vowels in the alphabet. In addition to similarity, %ORM of the PA list and transfer paradigms were systematically manipulated. The VD task was presented under the ANT procedure while the PAL was administered by both the ANT and ST methods. In the analysis of PAL no main effect of MP was revealed, however, a two-way interaction of MP with %ORM and three-way interaction of MP X %ORM X similarity was obtained. As mentioned before, the two-way interaction reflected the superiority of ST over ANT method with 100% ORM and the reversed relation under 50% ORM. The three-way interaction indicated depressed performance of a single group (medium-ST-50% ORM). Battig and Brackett attributed the three-way interaction to errors in the sampling of the population and warned against its generalization. However, interactions of similarity and MP has been reported by other investigators (Behring and Zaffy, 1965; Goss and Nodine, 1965; Experiment 10).

Behring and Zaffy (1965) varied only response similarity, employing similar or nonsimilar adjectives as response terms and CVCs with 53% associative value (Glaze, 1928) as stimulus terms. Subjects learned two lists, one with similar responses and the other with nonsimilar responses and each either under the ANT or ST methods. This procedure resulted in four conditions of (1) similar ST, nonsimilar ANT; (2) nonsimilar ANT, nonsimilar ST; (3) nonsimilar ST, similar ANT; and (4) similar ANT, nonsimilar ST. The main effect of MP reflecting superiority of ST method was found significant. Although the interaction of MP and similarity was not reliable in an overall analysis of variance, separate analysis of each MP revealed significant superiority of nonsimilar to similar conditions with the ST procedure but no difference due to similarity with the ANT method. These results indicate that the ST method is more susceptible to similarity variations than the ANT procedure and are opposite the prediction of the competition hypothesis.

The results are also in contrast with the findings of Goss and Nodine (1965; Experiment 10). Goss and Nodine employed a mixed list containing all four subsets resulting from factorial manipulations of two levels of stimulus and two levels of response similarity. The H similarity items were constructed of CVCs with two letters in common while the L similarity items shared only a vowel. Although no significant main effect of MP was obtained, its interaction with similarity was reliable and reflected a more pronounced effect of response similarity under ANT than the ST method. Moreover, the number of trials to criterion of one perfect trial for the L-L pairs was less for subjects under ANT procedure while the same measure for H-H pairs indicated superior performance of subjects under ST method. These findings are clearly in support of the interference hypothesis.

Several factors may be responsible for the contrast in the results of Behring and Zaffy and those of Goss and Nodine. First, is the difference in the type of similarity used in the two experiments. Behring and Zaffy employed associative or semantic similarity while Goss and Nodine manipulated formal similarity. It is possible, though perhaps unlikely, that the different types of similarity have opposite effects on MP. It is more likely that the contrast in the results stems from differences in the list design in the two experiments. Behring and Zaffy employed an unmixed list design while Goss and Nodine used a mixed list design. It may be remembered that list design produced differential effects of meaningfulness on MP. Demonstrations of differential effects of similarity on MP with mixed and unmixed list designs are also available. Goss and Nodine (1965; Experiment 9) reported an experiment identical to their tenth experiment in all respects except the employment of an unmixed list. The only significant effects obtained were produced by stimulus and response similarity.

Neither the main effect of MP nor its interactions with similarity were reliable. Although these results differ from those obtained by Behring and Zaffy in that superiority of L similarity to H similarity conditions were observed in both MP, it will be remembered that the interaction of response similarity and MP did not reach an accepted level of significance in the Behring and Zaffy study and the difference was obtained by separate analyses for each MP. No such analysis is reported by Goss and Nodine, but the mean number of trials to criterion measures indicate a larger difference between the L and H response similarity groups under ST (differences of $\overline{X}$s = 7.9 trials) than under the ANT method (differences of $\overline{X}$s = 2.6 trials). These results are similar to those reported by Behring and Zaffy and in contrast to the competition hypothesis expectations.

The lack of a similarity effect on MP of PAL is also reported by Cofer et al. (1967; Experiment III). They employed two eight-item lists composed of CVCs. List 1 items were constructed from only four consonants and thus were high in interitem similarity. In list 2, 15 consonants were used in construction of the item. When H inter-stimulus similarity was involved list 1 items were used for the stimulus terms and list 2 items as response terms. The reverse was employed for H inter response similarity. Cofer et al. found only the main effect of sex and its interaction with stimulus and response similarity to be significant. The main effects of similarity and MP produced null results. The lack of a similarity effect obviously stems from confounding of stimulus and response similarity. Since the high inter response similarity list contained low similarity stimulus items, and highly similar stimulus were paired with response terms of low similarity, any beneficial effect of low similarity of one functional item (stimulus or response term) was cancelled by the adverse effect of H similarity on the other. The lack of an effect of similarity and its interaction with MP, therefore, is expected in lieu of the procedures employed.

Two studies investigating list competition on ST and ANT methods of PAL are reported. Although these studies employed variables different from the traditional concept of interitem similarity, the similar effect of list competition and interitem similarity makes the placement of these studies in this section fitting. Gasparikova (1972) used the Spence, Farber and McFann (1956) competitive list in three methods of presentation (ANT, ST & ST-AB). The Spence, Farber and McFann list is a mixed list consisting of eight competitional and four non-competitional pairs. Gasparikova, based on the interreference hypothesis associated with the ANT method, predicted an interaction of MP with type of pair, suggesting a greater superiority of the ST method with competitional pairs than with non-competitional pairs. Analysis of number of intrusions confirmed the prediction. The ST method produced significantly fewer intrusions than the other two methods and ST-AB resulted in fewer intrusions than ANT method when the competitional pairs were considered. However, no difference among the three MP were obtained with an analysis of non-competitional pairs. These results provide strong support for compeition hypothesis of MP.

Voss (1969) examined the ANT and ST methods in probabilistic associative learning. In the task each stimulus was paired with two different responses in different pairs. The proportion of number of items that each of the two response terms was paired with a given stimulus was manipulated as a variable. In the ANT method the subject was required to respond with the response term he believed most likely to be paired with the stimulus term on a given trial. While in the ST procedure the subject was required to give the response that was paired with the stimulus on the immediately previous study trial. Voss used six different proportions of occurrence of the two responses under each MP. These were 1.00-0.0 (always the same response); 0.90-0.10; 0.80-0.20; 0.70-0.30; 0.60-0.40; and 0.50-0.50. It should be noted that response competition increases as the proportion of occurrence decreases from 1.00-0.00 to 0.50-0.50. It can be expected then that the superiority of the ST method would increase with decreases in the proportion of occurrences of response terms. Analysis of the number of correct responses confirmed the expectation. The superiority of ST increased as the probability of occurrence of one of the responses decreased. But this interaction was limited to the early trials. The effect many also be due to the procedural differences in the methods. While the subject was left to the mercy of his own guesses in the ANT method, in the ST procedure the pairing of response term in the immediately previous study trial was provided to the subject. This previous knowledge of the specific pairing, undoubtedly, is of special benefit when the proportion of occurrence of the two response terms are identical or very close (e.g., 0.50-0.50; 0.60-0.40).

Investigations of interitem similarity with PRO and ANT methods are limited to two. In both studies a multiple choice PA task was administered. Hawker's (1964) design consisted of a 2 x 2 x 2 x 2 factorial with two MP, two levels of response similarity, two levels of stimulus similarity and two forms of pair (either figures for stimuli and adjectives for response terms or the reverse). Similarity of the figure items was rated by ten judges and tested in a pilot study. The H and L similarity adjectives were obtained from the Underwood and Goad (1951) and the Melton and Saifier(Hilgard, 1951) lists, respectively. The subjects were presented with seven cycles each consisting of two training and one test trials. The rate of presentation was unpaced. No main effects of MP or its interaction with stimulus or response similarity were obtained with an analysis of the total number of correct responses on the seven test trials. A significant interaction of MP and trials, however, indicated the superiority of PRO method in the early trials. The lack of an overall effect of MP and its interaction with similarity seems to be caused by the unpaces rate of presentation as an analysis of amount of time consumed in each test trials indicated. Analysis of this measure produced significant superiority of PRO over ANT and a three-way interaction of stimulus similarity, response similarity and MP. The interaction indicated that with low stimulus similarity, increasing response similarity accentuated the difference between PRO and ANT procedures, while with high stimulus similarity increasing response similarity produced negligible differences between the two MP.

However, a later experiment (Hawker, 1965b) failed to yield an interaction of response similarity with MP. In addition to MP and response similarity, number of response alternatives (two and four) was manipulated in a factorial design. Only nonsense figure-adjective pairs were used in this experiment. In all other respects the two experiments were similar. The only effect involving MP obtained in the analysis of the number of correct responses or the amount of time taken on each test trial was its interaction with blocks of trials. The interaction once again reflected the superiority of PRO method in the early trials of learning. The lack of interaction of MP and response similarity is puzzling especially since the stimulus similarity was kept low in all groups and again reflects the inconsistent effect of interitem similarity on MP of PAL.

Only one study (Kanak, Cole & Eckert, 1972; Experiment I) has examined the possible effects of interitem similarity on methods of presentations of VDL. Variation of similarity was based on associative similarity. Kanak et al. employed three groups under each of the two methods of ST and ANT. In one group, interitem associate right, each of the two right items in the list were associatively related while the wrong items were unrelated. In the second group, interitem associate wrong, the wrong items were associatively related and the right items unrelated. In the control group no associations existed between the items. The frequency theory of verbal discrimination predicts superiority of associate right and the inferiority of associate wrong to the control group through the mechanisms of implicit associative responses. Kanak et al. argued that ST presentation may facilitate evocation of implicit associative responses and may aid demonstration of the rarely obtained effect of interitem associations in VDL. In addition, intitial trial guessing or no guessing was used as a variable. The ST method of presentation produced significantly less errors than ANT in all three groups. Neither the main effect of interitem relation nor its interaction with MP were reliable.

As was the case in meaningfulness investigations, the inconsistent results of studies of interitem similarity does not permit any meaningful conclusion in regard to its effect on MP.

Intrapair Similarity. Cofer et al. (1967; Experiment II) are the only investigators examining the role of intrapair similarity on the MP of PAL. They employed three levels of similarity. All lists were comprised of CVCs and the stimulus terms were the same in all conditions while the response term differed for each level of similarity. For the high similarity group, the response terms shared two letters with the stimulus terms. For the medium groups only one letter was shared and for the low groups no letter was shared between stimulus and response terms. It should be remembered that the interference hypothesis predicts an increase in the difference between the methods of presentation in favor of ST with low intra-pair similarity and subsequently greater difficulty in the associative learning stage. This prediction was not confirmed. Superiority of the ST method was demonstrated across all levels of similarity as reflected in the significant main effect of methods. Although the advantage of ST presentation was greater under the medium similarity condition than the other two conditions, the difference did not reach a reliable level. Whether these results are reliable or not remains to be tested by further investigations.

Imagery. The comparative effect of imagery on the ST and ANT methods of PAL has not been systematically investigated. Kopstein and Roshal (1961), however, reported an experiment in which verbal or pictorial stimulus terms were factorially manipulated with a variation of the PRO and ANT methods. Insofar as pictures can elicit better images than words, this study can be considered as a rudimentary investigation of imagery with MP of PAL. The Kopstein and Roshal design consisted of a 2^4 factorial with two types of stimuli during training trials, two types of stimuli during test trials, two verbal contexts (words and connected discourse) and two MP. The methods were identical to those used in the Kopstein and Roshal (1955) study described earlier. The verbal stimuli were highly frequent English words and the picture stimuli were drawings of the verbal stimuli. Response terms were the Russian language equivalent of the stimulus terms. Performance was better with pictorial stimulus terms and with the PRO method, but the two variables did not interact.

The lack of interaction of imagery and MP is also reported with ST and ANT in VDL (Ingison & Ekstrand, 1970; Rowe & Paivio, 1972). In the earlier of the two investigations, as reviewed before, imagery and

MP were factorially manipulated with a number of other variables such as frequency and feedback presentations. A mixed list containing pairs of high and low frequency and imagery was employed. Concreteness and abstractness measures were selected from Gorman's (1961) compiled list scales. Neither imagery or frequency affected performance of VD learning in any way.

Rowe and Paivio (1972) also used a mixed list to investigate the effects of imagery on the ANT and ST methods of VDL. The H imagery items had an average value of 6.38 versus an average of 3.28 for L imagery items, based on the Paivio, Yuille and Madigan (1968) norms. In contrast to the results of Ingison and Ekstrand, the main effect of imagery was significant reflecting a facilitative effect of high imagery on VDL, while the main effect of MP was not reliable. However, as in the former study, no interaction of imagery with method was observed. Although Rowe and Paivio could not identify the course of discrepancy between the two experiments, a number of procedural differences (i.e., list length, rate of presentation), could be responsible for the differential outcome, but it is more likely that the source of conflicting results lies in the difference normative scales of imagery used in the two studies and perhaps the extremes of imagery values sampled.

Investigations of the effects of attributes of materials on methods have provided little insight into the sources of inconsistent results. Manipulations of stimulus and/or response term meaningfulness and inter-item similarity produced inconclusive results. However, two main trends are apparent in the effects that these variables produced. First, the superiority of the ST method and interaction of methods with meaningfulness and similarity are more potent when a mixed list design is used for manipulation of these attributes. Secondly, the interaction of methods and material attributes, when obtained, supports the interference hypothesis. The effect of intrapair similarity on methods of presentation is limited to the results of one study and any conclusions in this regard may prove to be premature. Insofar as the results of the three studies involving imagery can be generalized, this variable seems to have no specific interaction with methods of presentation.

Rate of Presentation

The rate of presentation can be manipulated in two basic manners. One method consists of varying the time allotted for anticipation (or test) and/or feedback (or study) periods. In the other methods, used primarily in the investigation of massed versus distributed practice, either the intertrial interval or the interval between the presentations of items is manipulated. It is generally demonstrated that slower rates of anticipation or feedback presentation results in superior performance in terms of number of trials to criterion or total errors. On the other hand, in a review of investigations of massed and distributed practice using the ANT method of PAL, Underwood and Ekstrand (1967) found no consistent effect of the intertrial interval manipulations on performance. The effect of rate or anticipation or feedback presentation is undoubtedly due to the time allotted for processes of storage and retrieval. A longer feedback period results in superior storage and a longer anticipation period produces better retrieval of the material. It can be reasonably assumed that the effect of rate of presentation of feedback and anticipation periods is accentuated in task involving interference, since such task requires more complex storage and retrieval mechanisms. Based on this assumption, the competition hypothesis of Kanak and Neuner predicts a greater difference between the ST and ANT methods with faster rates of presentation than with slower rates. Based on the same argument it is expected that the rate of presentation would have a greater influence on the ANT method. As Izawa (1970; 1971) has pointed out, this reasoning can be extended to the effects of the intertrial interval. In the ANT method the intertrial interval has no effect since this procedure involves an inherent massed practice of storage and retrieval in each trial. However, in the ST method, where the two processes are separated, an increase in the intertrial interval is expected to facilitate learning.

Anticipation and Feedback Rate of Presentation. The three investigations of rate of presentation in ST and ANT procedures reveal no differential effect of the rate variable on performance under the two MP. Cofer et al. (1967; Experiment I) covaried rates of presentation of feedback and anticipation periods in a PA task. The rate of presentation were 5:5, 3:3 and 1:1 sec with a 30 sec intertrial interval constant in all conditions. The mean number of trials to criterion for the 1:1, 3:3 and 5:5 rates were, respectively, 32.75, 15.00 and 9.18 under ANT method and 23.6, 8.06 and 5.69 under ST procedure. The overall superiority of ST was reliable over all the rate conditions. The expected increase in superiority of ST with faster rate was not obtained as reflected in the nonsignificant interaction of the two variables.

Similar results were obtained in now-familiar Ingison and Ekstrand (1970) study where the feedback rate was manipulated in a VDL task. The frequency theory predicts an increase in learning proficiency with a longer feedback period. This prediction is based on the assumption that the subject's rehearsal of the right items increases with longer feedback period, resulting in greater frequency accrual to the right item. Three rates of feedback presentation (2, 3 & 6 sec) were employed. Although the rate of presentation affected VD performance as predicted by the frequency theory, and superiority of ST over ANT method reached significant level, the interaction of the two variables was not reliable, thus, extending the Cofer et al. findings to VDL.

Battig et al. (1963) manipulated the rate of test trial presentation in the ST method of SL. The two different rates of test trials (4 and 8 sec) had no effect on learning efficiency. This finding is in line with the expectation of the interference hypothesis that the effect of rate of presentation on ST is less pronounced than its effect on ANT method, but the lack of comparable anticipation conditions prevents any meaningful conclusion. Based on the results of the former two studies, however, it seems clear that the rate of presentation does not play a major role on the outcome of comparisons of the ST and ANT methods and inconsistent results in the literature cannot be explained by rate of presentation of feedback and anticipation periods alone.

Massed and Distributed Practice. The effect of manipulations of both the intertrial interval (ITI) and interitem interval (I-item-I) on ANT and ST methods has been extensively investigated by Izawa (1970, 1971). Izawa (1970) reported two experiments in which either ITI (Experiment I) or I-item-I (Experiment II) were manipulated in a ST procedure of PAL. The four conditions of ITI employed included 0, 24, 48 and 72 sec intervals between stury-test as well as test-study trials. Likewise, four I-item-I conditions (0, 2, 4 & 6 sec) were included in the second experiment. Since the list consisted of 12 pairs, the total time of each ITI condition corresponded with that in the appropriate I-item-I condition. The intervals were filled with a neutral task of naming colored geometrical figures. The results demonstrated a pronounced effect of distributed practice as increases in ITI or I-item-I reduced the proportions of error.

A more extensive examination was reported in the 1971 paper. In a series of nine experiments Izawa namipulated ITI and I-item-I in ANT or ST method (though not factorially) and in study or test trials of ST using a neutral task, naming of colored geometrial figures (N), or blank trials (B) to fill the intervals. Table 3 presents the design of the nine experiments.

The first two experiments, in both of which the ANT procedure was employed, differed from each other only in terms of the locus of the interval (ITI or I-item-I). The I-item-I was always given after the feedback period. The remaining seven experiments were devoted to investigation of masses and distributed practice in the ST procedure and differed in terms of locus of interval, whether the I-item-I was given during study, test, or both study and test trials. The interval in all experiments was filled with the neutral task with the exception of the last experiment in which blank trials were used as interval fillers. Six ITI conditions (o, 24, 48, 72, 96, & 144 sec) and six corresponding I-item-I conditions (o, 2, 4, 6, 8, & 12 sec) were employed in the first two experiments, while the number of interval conditions were reduced to the first four in all other experiments.

In neither of the first two experiments were any significant differences between the conditions observed. In fact, in the second experiment, in which effect of I-item-I was studied, best performance was obtained in the massed practice (o sec I-item-I) condition. These results confirm the previous findings of lack of an ITI effect on the ANT procedure of PA learning.

The facilitative effect of distributed practice, however, was obtained in Experiments III and IV which involved ITI or I-item-I, respectively, after or during study trials of the ST method. The massed practice conditions in both experiments were significantly inferior to their corresponding distributed practice conditions which did not differ from each other. In comparisons of the two experiments no difference between I-item-I and ITI conditions was observed in the early trials, but superiority of I-item-I over ITI conditions was evident in later trials.

The effects of distributed practice was not pronounced in the experiment investigating its effects on test trials of ST procedure (Experiments V & VI). Although Izawa tends to present the results in favor of superiority of distributed over the massed conditions by the use of multiple t-tests, p values for the overall analysis of variance on both experiments were never less than .10 and indeed, according to her own graphs, one distributed condition (24 sec ITI) in Experiment V produced inferior learning when compared to the massed practice condition.

In contrast, the results of the last three experiments, in all of which intervals were placed after or during both study and test trials, revealed a pronounced facilitative effect of distribution of practice. Differences in performance under the different conditions of Experiments VII and VIII were substantial and much greater than was observed in Experiments III and IV or V and VI. Best separation among the conditions was obtained in the last experiment in which blank trials were employed during the interval periods. Comparisons of Experiments VII, VIII and IX revealed significantly more potent effects in the last experiment.

Several conclusions can be drawn from the results of the nine experiments. First, distributed practice does facilitate PA learning under ST method but it has no such effect under ANT procedure. Secondly, the distributed practice effect is more potent when intervals are placed after or during study trials than when they are placed in test trials. Thirdly, although the effect of introduction of intervals during or after test trials are negligible when combined with intervals during or after study trials it enhances the facilitative effect of distributed practice. Finally, the effect of distributed practice is more potent when blank trials are employed during the intervals than when a neutral task is presented, indicating rehearsal of items occurs during the interval when no other task is presented. These results may be expected to renew research interest in the classical problem of massed versus distributed practice on verbal learning tasks, an issue all but buried after publication of Underwood's (1961) exhaustive examination of the topic.

As mentioned earlier, these results prompted Battig (1973) to attribute inconsistencies in superiority of ST method to variations in ITI employed in different experiments. Battig pointed out that Battig and Brackett (1961) and Cofer et al. (1967; Experiment I) studies, in both of which a clear advantage of the ST method was observed, were unique in that they employed a relatively long ITI of 30 sec. Although variations in ITI may contribute to discrepancies in the results, a closer examination of the experiments involving comparison of ANT and ST procedures reveals that ITI variation is not the sole determinant to the inconsistent effects. For example, Battig et al. (1963; Experiment I) and Cofer et al. (1967; Experiment VI) used ITIs of 30 and 18 sec, respectively, and found no difference in PAL under the two methods. On the other hand, numerous experiments across different tasks (Battig & Wu, 1965; Cofer et al., 1967, Experiment II; Izawa, 1972, 1974; Kanak & Neuner, 1970; Voss, 1969; Battig & Switalski, 1963; Fulkerson & Johnson, 1971; Posnansky, 1967; to list a few) have all employed ITIs of 6 sec or less and all reported significant superiority of ST over the ANT method.

List Length

It will be remembered that Izawa's retention interval hypothesis predicts a greater superiority of the ST method with longer lists. This prediction is based on the assumption that the greater lengths of list produces increasing differences between the retention intervals of given pairs for the ANT and ST procedures. Izawa (1974) pointed out that since greater retention intervals are assumed to produce poorer performance by losses in short-term memory, the superiority of ST method is expected to diminish with very long lists as the retention interval for the ST procedure, as well as for the ANT method, will exceed the boudnaries of the short-term memory span.

Accordingly, Izawa (1974) examined the effects of list length on performance of PA learning with ANT and ST procedures in two separate experiments. In the first experiment four list lengths (5, 10, 15 & 20 pairs) were employed while in the second experiment only two list lengths (30 & 40 pairs) were used. In all other respects the two experiments were identical. Izawa predicted no difference among the two methods with list lengths on the two extremes of the variable, but superiority of ST method with medium list lengths. The results of the two experiments confirmed her predictions. Both main effects of MP and list length were significant. In addition, the interaction of the two variables approached a reliable level. Superiority of ST over the ANT procedure was reliable with the 15 and 20 pairs lists, approached significance with the 30 pair lists, but was not obtained with the 5, 10 and 40 pair lists.

Similar results were obtained with a VD task. Underwood, Shaughnessy and Zimmerman (1972) investigated the effects of list length, MP and degrees of interpolated learning on VD performance. The inclusion of list length as a variable primarily served the function of testing the frequency theory of VD which assumes no difference in performance with different list lengths. In addition, it was argued that the list length manipulation can aid in determining some of the mechanisms involved in MP. If delay of feedback has any influence on VDL by ST method, the effect should be accentuated with a long list. They stated that it is possible for the ANT method to yield better performance with a long list. On the other hand, if the separation of learning and performance is the critical factor, length of list should be irrelevant. Two list lengths (15 & 45 pairs), two methods (ST & ANT) and two degrees of interpolated learning (3 & 9 trials) constituted the factorial design of the experiment. Contrary to the Underwood et al. expectations, but in support of the retention interval hypothesis, the greater length of list resulted in greater superiority of ST over the ANT method. The performance on the 15 pair list under the two MP was virtually identical, while performance under ST was consistently, and significantly, superior to that under the ANT method with the 45 pair list.

Goss and Nodine (1965; Experiments 9 & 10) employed length of list manipulations to investigate whole versus part list learning under the ST and ANT methods of PAL. Subjects learned a 12-pair mixed list either as a whole or in parts of 4 pairs. The lists were mixed in terms of stimulus or response meaningfulness (Experiment 9) or similarity (Experiment 10). No difference with respect to list length was observed in performance under the two MP. Under both lengths, superiority of the ST method was obtained.

Cofer et al. (1967), in examining the results of Goss and Nodine, found that an advantage of ST was observed with a mixed list but not with an unmixed list, argued the effect may be due to the possibility that in the 12-paired list subjects could learn the list as though it were composed of several shorter lists. Cofer et al. thus concluded list length may be a factor influencing the outcome of MP comparisons and predicted a greater superiority of the ST method with shorter lists. However, tests of their hypothesis in two experiments (Experiments IV & V), which were identical with the exception of length of lists (10 and 4 pair, respectively), produced results opposite their expectations. Although the ST method was slightly superior, yet nonsignificantly so, with the 10 pair lists, the ANT method produced better performance, and significantly so with one dependent variable, with the the 4 pair list. Such results obviously support the retention interval hypothesis.

Izawa (1974) suggested list length to be the major source of inconsistency in results of comparisons of MP. Although the few studies that have varied list length under both ST and ANT have generally produced results in support of this notion, list length alone cannot explain the inconclusive results. Goss and Nodine did find superiority of the ST method with a very short mixed list. Moreover, Cofer et al. (1967; Experiment VI) in a replication of their first study in the series in which superiority of the ST procedure was demonstrated over all three rates of presentation, employed a longer list and found no difference in performance between the two MP. It should be noted, however, that the two studies differed also in ITI with the first study having a longer ITI than the sixth, a variable which may have contributed to the differences in the outcome of the two experiments. In addition, a closer examination of the literature reveals a number of experiments which have employed lists with 8 pairs, or less and have reported superiority of ST over the ANT method (Behring & Zaffy, 1965; Horton & Wiley, 1967; Kanak & Neuner, 1970; Kausler, 1963; Voss, 1963, Wright, 1967) while other investigators used list lengths of 8 to 16 pairs and found null differences in performance under the two MP (Battig _et al._, 1963; Cofer _et al._, 1967, Experiments III, IV, VI; Lockhead, 1962; Reynolds, 1969). It is thus apparent that list length may be a contributing factor to the inconsistent results, but it cannot be the sole contributing factor.

Personality Variables

The competition hypothesis predicts an interaction of some personality variables with MP. One such variable, ego strength, has already been discussed in detail in an earlier section of this paper. Duffy and Kanak (in press) predicted, and demonstrated, that high ego strength subjects performed better under the ST method while low ego strength subjects learned somewhat better under the ANT method of PA learning. Another variable which is expected to interact with methods of presentation, according to the interference hypothesis, is anxiety level. The presence of an intrinsic source of competition in the rapid alternation of storage and retrieval mechanisms in the ANT method and lack of this source of competitiontion in the ST procedure is the basic assumption on which the prediction of an even greater superiority of ST method with high anxiety subjects than with low anxiety subjects, is based. Unfortunately, the only test of interference hypothesis interpretation with regard to personality factors is that reported by Duffy and Kanak. The expected interaction of anxiety with MP awaits empirical confirmation.

A number of investigations have extended investigation of MP comparisons to populations other than college students. Two studies have been reported with retarded children. Blue (1970) compared performance of normal and retarded adolescents under ST and ANT methods of PAL. A ten pair list was employed with geometric designs as stimulus terms and color names as responses. The normal subjects had a mean IQ of 103 and the retarded subjects had a mean IQ of 70. The MP was introduced as a within-subject variable as subjects learned two lists, one under each procedure. Analysis of trials to criterion revealed that retarded subjects performed better under the ST method, while the performance of normal subjects was unaffected by MP. It should be noted, however, that although Blue does not make a note of it, an examination of the two lists used in the experiments for first and second list learning, reveals an A-B, A-C paradigmatic relationship existed between them and the results of the experiment should be considered with caution since this paradigm should also produce an interaction with the two methods, according to the interference hypothesis.

Stolurow and Lippert (1964) investigated PRO and ANT methods of PA learning with retarded children. Two levels of practice were used. Subjects reached the lower criterion of learning significantly faster under the PRO method, but little difference between the two MP were obtained when the learning criterion was increased. These results are similar to those obtained with college students (e.g., Hawker, 1965a) indicating the superiority of PRO method to be limited to early trials.

An extension of MP effects to a geriatric population is reported by Kausler (1963). Subjects were geriatric non-psychiatric patients matched in age and level of education. They were presented a mixed list consisting of low and high intrapair associative pairs under the ST or the ANT method of PAL. Analysis was performed only on the low associative pairs since the high associative pairs were learned very fast in both groups. Although the main effect of MP was not found reliable, a significant interaction of blocks of trials and MP indicated superiority of the ST method in the last of the two blocks of three trials. The results of these three investigations seem to point to the conclusion that the superiority of the ST and PRO methods to the ANT method is generalized to populations other than college students.

Pronounciation of Items

It should be remembered that Cook and Kendler (1956) attributed the advantage of the PRO procedure, at least in part, to shorter between-terms delay, the omission of overt responses, or both. Likewise, Cook and Spitzer (1960) argued that the occurrence of overt practice during the anticipation interval would interfere with the establishment of the essential stimulus-response connection that must take place over the interval and may also interfere with learning. It should be noted this argument differs from that advanced by Peterson and Brewer (1963) who attributed the inferiority of the ANT method to the interference stemming from wrong responses. The covert and overt pronounciations of the items, therefore, is a variable of theoretical importance in comparisons of the PRO and ANT methods. However, this variable has little theoretical relevance in investigations of the ST and ANT procedures s.nce none of the three major formulations allots any role to pronounciation in their explanation of the superiority of ST over ANT method.

To test their hypothesis, Cook and Spitzer (1960) designed a two by two factorial experiment with two MP (PRO & ANT) and two modes of practice (no pronounciation, N and overt pronounciation, O), during anticipation intervals of training trials. They predicted the PRO-N condition to produce the best and the ANT-O to result in the worst performance. SUbjects were presented 27 training trials with a test trial after every three training trials. No factorial analysis was reported, but an analysis by the Mann Whitney U test on the number of correct responses on each of the nine test trials produced the following results. The PRO groups (across both levels of pronounciation) produced significantly more correct responses in the first six test trials. Likewise the N groups (across MP) were superior to the O groups in the first six test trials. These results indicate that both "delay" and overt pronounciation contributed to the inferiority of ANT procedure. Moreover, while the superiority of PRO-N condition over the ANT-N condition was evident only in the early learning, the superiority of the PRO-O condition over the ANT-O was consistently significant up to the eight test trials suggesting that overt pronounciation has a substantially greater adverse effect on ANT than it has on the PRO method. This effect may be due to the lack of incorrect responses in the PRO method to interfere with associative learning. The effect of pronounciation on MP was also studied by Reynolds (1967) in an a forementioned experiment. In addition to the MP and pronounciation variables, Reynolds varied the number of test trials. Half of the subjects received a test trial after the third, sixth and ninth training trials while the other half had nine uninterrupted training trials followed by a single test trial. Although the main effect of pronounciation mode was found significant in favor of covert pronounciation conditions, its interaction with MP was not reliable, failing to replicate Cook and Spitzer's findings. It seems apparent that overt pronounciation during the anticipation interval hinders performance. Whether this factor has a major role in producing the superiority of PRO remains to be answered by further investigations.

The investigation of the effects of pronounciation of items on ST and ANT procedures of PAL is limited to one study. Schild and Battig (1966) examined bidirectional and unidirectional PA learning under both methods. Half of the subjects pronounced both stimulus and response terms once and the other half twice during the feedback, or study, interval. Pronounciation of items did not effect the performance in any way. Subjects under both pronounciation conditions learned faster with the ST procedure than with ANT.

The effects of pronounciation on ST and ANT methods of VDL has been examined by Fulkerson and Johnson (1971) and Rowe and Paivio (1972). In the Fulkerson and Johnson study the initial trial guessing was also varied. Subjects were instructed either to guess the right items on the first trial or merely observe the items. In the ST method the first trial was a test trial under the guessing condition. The frequency theory of VDL predicts the no guessing condition to produce better performance as the greater pronounciation of the wrong items under guessing would increase the frequency units of the wrong items. Analysis of error measures produced an interaction of MP and the initial trial practice. Subjects in the ANT guessing condition made more errors than subjects in all the other three conditions, which did not differ from each other. However, no such interaction was observed when Rowe and Paivio instructed the subjects to either pronounce both wrong and right items during the feedback period or not to pronounce them, even though pronounciation of the item resulted in inferior performance as the frequency theory predicts.

The effect of pronounciation on MP is inconsistent, however, there seems to be a weak tendency for ANT method to be affected to a greater degree by pronounciation of items than the ST procedure. This may be caused by the accentuating effect that the requirement of overt pronounciation may have on an already existing intrinsic source of interference between storage and retrieval mechanisms in the ANT method.

Serial Order of Presentation

A number of investigators have tested the effects of constant and varied order of presentation on MP. This variable has no theoretical importance in terms of MP as none of the major hypotheses predicts it to have a differential effect on MP. In general serial order of presentation is employed as a variable to investigate theories unrelated to MP and the inclusion of methods of presentation as a variable serves only as a test of generality of the specific findings.

Battig et al. (1963) investigated the effect of constant and varied serial order of presentation of PAL, in a PA to SL transfer task, in a series of five experiments, two of which (Experiments I & IV) also included different MP. In the first experiment three MP were employed consisting of ANT, ST and a variation of ST procedure, ST-5, in which five study trials preceded each test trial. In the fourth experiment only ANT and ST methods were used. The items used in the SL lists were either the same as the response terms, the same as the stimulus terms or had no relations to the PA lists. It was predicted that a constant serial order in PAL would result in greater positive transfer to SL than varied order. The results of Experiment I were ambiguous as the sequence of order of varied or constant list learning (a within-subject variable) interacted with methods. Analysis of the first PA list learning resulted in an interaction of the serial order of presentation and MP, indicating the constant order of presentation reduced errors only in the ANT conditions. However, this interaction was not obtained in the fourth experiment, a result which led Battig et al. to attribute the effects found in the first experiment to the nonequality in ability of subjects in different groups.

The lack of interaction of MP and constant versus varied serial order of presentation is also reported by Battig and Lawrence (1967) in an SL to SL transfer task. Similarly constant or random serial order of presentation of the items (Fulkerson and Johnson, 19/1) and constant or varied spatial position of items (Rowe & Paivio, 1972) produced no specific effect of performance under the ANT or ST methods of VDL. Finally, Cavanagh and Parkman (1971) reported superiority of ST method over ANT in a procedure, called "steady state" by the authors, in which each item correctly recalled was replaced by a new item. Although traditional ST and ANT methods were not included to determine possible differential effects of steady state presentation on MP, the results do demonstrate an extension of superiority of the ST method with a different methodology.

It seems clear that the serial order of presentation has little, if any,interaction with MP and its exclusion from the attention of the major theoretical positions is well justified.

Transfer of Training Tasks

Although transfer mechanisms have been investigated extensively, the interest in the effects of MP on transfer of verbal tasks has been limited. The theoretical importance of transfer for comparisons of methods of presentation stems primarily from the associative interference produced in PAL negative transfer paradigms. The interference theory of Kanak and Neuner predicts that the addition of intrinsic sources of interference, such as in negative transfer paradigms will combine with the competition intrinsic to the ANT method to produce even greater inferiority to the ST method. Thus, if both positive and negative transfer paradigms were studied in combination with the ANT and ST methods, the ANT method would be expected to yield poorer transfer performance in both types of paradigms, with the relative inferiority to the ST method being greater in negative transfer paradigms. Unfortunately, no systematic investigation of these variables has been attempted and the predicted interaction, when examined, has been rarely obtained.

Peacock (1971) examined PAL transfer under the ANT and ST-AB methods, the latter not being the traditional ST method; but one requiring recall of both A and B terms on test trials. Four transfer paradigms were employed: A-B, A-C; A-B, C-B; A-B, A-Br, and the nonspecific control paradigm of A-B, C-D. Since the ST-AB method produces symmetrical bidirectional associative learning and the ANT method produces asymmetrical (i.e., S-R R-S) bidirectional leanting (Kanak & Neuner, 1970), Peacock predicted that the ST-AB method would produce more negative transfer in the A-B, A-Br and A-B, C-B paradigms than the ANT procedure. This prediction was not confirmed. No interaction of methods and transfer paradigms was obtained, although the overall superiority of ST-AB over the ANT method and the control over the A-B, A-C and A-B, A-Br paradigms was found significant. The lack of interaction between the two variables may be due to two contrasting mechanisms. On one hand the symmetrical associative learning in list one under ST-AB increases the negative effect of backward or B-A associations in the A-B, A-Br and A-B, C-B paradigms. On the other hand, the intrinsic interference between storage ans retrieval mechanisms in the ANT method makes it more susceptible to the effects of negative transfer paradigms. These two mechanisms could produce counteracting functions resulting in the lack of interaction. It should be noted, however, that in the A-B, A-C paradigm, in which only unidirectional A-B associations interfere and which are assumed to be of equal strength in the two methods, an increased superiority of ST-AB over ANT is predicted, when compared with the control paradigm, by the interference hypothesis of MP. No such effect was observed, however, the ST-AB method requires more learning qua learning than the ANT method by virtue of requiring retrieval or recall of two terms rather than one. A more sensitive test of the proposed interaction would therefore be made is the traditional ST method were compared with ANT.

Horton and Wiley (1967) reported results partially supporting the interference hypothesis. They investigated two three-stage mediation paradigms under the ST and ANT methods. The mediation paradigms consisted of a forward chaining paradigm (A-B, B-C, A-C) and its corresponding control paradigm (A-B, D-C, A-C). Horton and Wiley reasoned that the employment of a mediation design makes it possible to evaluate the relative superiority of ST method under varying conditions of stimulus learning, response learning, and associative strength. Performance on the second stage of the mediation paradigm provides an evaluation of the effect of stimulus learning and the third stage performance provides an opportunity for evaluation of response learning and associative strength effects. In terms of the interference hypothesis, the second stage in the experimental condition produces backward associative interference and, thus, it is expected that superiority of ST over the ANT procedure would be enhanced on that stage. This prediction was confirmed. The mean number of correct responses in the first five trials of the first, second and third stages were, respectively 16.12, 19.29 and 15.50 under the ANT procedure and 23.42, 26.19 and 13.42 under the ST method in the experimental condition. The superiority of ST over ANT method in the first two stages was significant while the superiority of ANT method on the last stage was not. The lack of difference between the two methods at the third stage may be due to learning to learn effects, as the subjects may develop some strategy to inhibit the interference effect inherent in the ANT method.

Inquiries into the effects of MP on SL transfer have been reported by Battig and Lawrence (1967) and Posnansky (1972). Battig and Lawrence examined two transfer paradigms. In one paradigm, reversal (REV), the second list was the exact mirror image order of the first list, while in the other paradigm, scrambled (SCR), the second list consisted of an unsystematic rearrangement of the first. These paradigms are generally employed to examine positional and sequential hypotheses of SL. The sequential hypothesis predicts positive transfer with REV and negative transfer with the SCR paradigm, while according to the positional hypothesis both paradigms should produce negative transfer. In addition, two modes of serial ordering were used factorially in the first and second lists. In the constant mode condition the list began and ended with the same items across all trials. In the varied conditions the beginning of the lists rotated across trials although the sequential serial relation of items remained constant. It should be noted that in the varied conditions the positional cues of items are changed on each trial, thus, according to the positional hypothesis, learning would be more difficult in those conditions than in the constant conditions. Performance on both lists was better under ST than the ANT method and with a constant serial order. In addition, analysis of List 2 performance resulted in superiority of the REV over the SCR paradigm. More important was the finding that the ST procedure accentuated both transfer and serial position mode effects. The mean number of trials to criterion of the second list in the REV paradigms were 8.0 and 2.6, respectively, for the ANT and ST methods. The same measures for SCR paradigms were 11.2 and 8.3. In contrast to the prediction of the interference hypothesis of MP, the negative transfer paradigm reduced the superiority of ST over ANT method. These findings also reflect a greater sensitivity of ST method to paradigmatic influences. Battig and Lawrence argued that the sensitivity of ST procedure stems from reduced confusion of subjects in this MP permitting more effective operation of more complex multiple cue (positional and sequential) processes as evident in near perfect performance of subjects in the ST condition under REV paradigm, when both first and second lists were administered under the same serial mode of presentation.

Posnansky (1972), in part, replicated the Battig and Lawrence experiment by the employment of the REV and SCR paradigms under the two MP of ANT and ST. In addition, grouping of the items and probing cues were varied, the latter as a within-subject variable. Items were either presented one after another or were separated in groups of three by a blank space. Three probing methods were employed. These were presentation of the numerical position of the item cue, 3-prior-items cue and prior-item cue. The superiority of ST over the ANT method was obtained in both lists and the REV paradigm produced superior performance. Moreover, interactions of MP with probing and serial positions were found to be significant. Subjects under ANT method used 3-prior-item cues effectively for more items than did subjects under the ST procedure. On the other hand, ST subjects used prior-item cues more effectively. In addition, interaction of MP and serial order position with analysis of 3-prior-item cue suggested that the 3-prior-item cue was least effective under ANT procedure at the very center and the extreme positions while it was most effective at the intermediate positions. In contrast, under ST the 3-prior-item cue usage was greatest at the beginning serial positions with its usage being attenuated across subsequent positions. These results may indicate substantial differences in the mechanisms of SL under the two MP. It seems apparent that the ANT procedure requires greater cues for an effective learning of a SL task. Although Posnansky replicated the transfer results of Battig and Lawrence in that REV paradigms produced superior learning, no interaction of MP and paradigm was observed, indicating, once again, an inconsistent effect of MP.

Battig and Brackett (1963) have investigated transfer of VD to PAL. As mentioned before, Battig and Brackett employed intralist similarity and % ORM as well as transfer paradigm variables in their examination of the ST and ANT methods. Seven different transfer paradigms were presented in a mixed-list design. In the DISC (S) subsets the stimuli of the PA list consisted of the right items and the responses of the PA list were the same as the wrong items of the VD list. The DISC (R) subset was just the reverse of DISC (S). In the FAM (R-S) subset both stimulus and response terms in the PA list were derived from the right items of VD list whose wrong items did not appear in the PA list. In FAM (O) subset the stimulus and response terms of the PA list were constructed from the wrong items of the VD list. The FAM (S) subset consisted of the right item of the VD pairs whose wrong item did not appear in the PA list, as the stimulus term and the wrong items of another VD pair as the response term. The FAM (R) subset had the revered relation of FAM (S). There were two pairs from each the DISC conditions, one pair from each of the FAM subsets and four pairs constructed from new items in the PA list. Battig and Brackett predicted superiority of DISC pairs over the FAM pairs. The predictions, based on previous results of similar experiment (Battig, Williams and Williams, 1962), stem from the fact that the DISC pairs are analogous to the A-B, A-B or A-B, B-A paradigms, while the various FAM pairs are analogues of the A-B, A-C; A-B, C-A; A-B, B-C; or A-B, C-B paradigms and the new pairs form an A-B, C-D paradigmatic relationship. As predicted, PA performance was slightly better in DISC than new subsets and poorest in the FAM pairs, the difference being significant only in comparisons of DISC and FAM subsets. A more interesting finding was the interaction of MP with types of pairs. The differences between subsets were highly significant under the ANT procedure while no differences were observed under the ST method. Their findings clearly support the interference hypothesis of MP.

Retention.

The retention interval hypothesis predicts that the superiority of ST method during acquisition diminishes in retention measures if the recall test is given after a relatively long retention interval. This argument is based on the assumption that the advantage of ST procedure during acquisition is caused by the aid of short-term memory processes with shorter intervals between feedback and test periods for a given pair in that procedure. As the retention interval increases, it exceeds the span of short-term memory and, thus, it cancels the superior acquisition effects of the ST procedure. In a test of this hypothesis, Izawa (1972; Experiment II) examined the recall of PAL following a 15 minute retention interval in the ST and ANT methods. The recall trial was identical to acquisition trials. No difference was obtained in the recall measures of the two methods, as predicted by the retention interval hypothesis, in spite of the significant superiority of ST procedure in acquisition performance. Unfortunately, no other studies have examined retention under the two MP with long retention intervals. A number of investigators, however, have reported immediate retention tests of stimulus or response terms with different MP. Since the immediacy of recall following acquisition makes it possible for short-term memory processes to be still functioning, these studies cannot be considered as a test of the retnetion interval hypothesis. Nonetheless, superiority of the ST method when manifested during acquisition has seldom been reported to carry on in even immediate recall measures.

Schild and Battig (1966) reported administration of two recall tests to subjects. In one test all the items of the PA list (both stimulus and response terms) were presented singly to the subject who was asked to recall the paired member of each item. On the second test subjects were asked to free recall as many pairs as they could. No difference was found in recall of items or pairs under the ST and ANT methods, which were also factorially combined with unidirectional versus bidirectional practice, even though the superiority of ST during acquisition on both types of directional practice was established. This lack of differential effect, however, may be due to the unpaced nature of the recall tests.

Similar results were reported by Kanak and Neuner (197). It will be remembered that these investigators examined learning of backward and forward associations under ANT, ST-B, ST-A, ST-AB and REC conditions and found that ANT and ST-AB methods resulted in significantly slower rates of acqusition than the other three conditions. Following acquisition, half of the subjects were given an unpaced stimulus-response recall test and the other half an unpaced response-stimulus recall test. A second recall test, counterbalancing the order of the directional tests was administered after completion of the first recall test. Although recall under the different method differed in terms of directionality, no difference in backward or forward recall was observed in comparisons of ANT and its counterpart, the ST-B method.

Other investigators have reported interactions of MP with word attribute variables in analyses of backward recall measures. Wright (1967) varied stimulus meaningfulness in mixed and unmixed lists. Superiority of ST method acquisition was obtained under the mixed list but not under the unmixed list. Recall data of the paced backward recall test revealed an interaction of stimulus meaningfulness and MP. More omissions in recall resulted under ANT than the ST method, furthermore, these differences were greater under low stimulus than high stimulus meaningfulness. Interaction of the meaningfulness of items and MP in recall measures is also reported by Goss and Nodine (1965; Experiments 7 & 11). Goss and Nodine varied stimulus and response terms meaningfulness factorially in an unmixed list design. In both studies, which were replications, a paced backward recall test was given immediately following termination of acquisition. Although no main effect or interaction due to MP was found significant in the analysis of acquisition data, analysis of the recall data revealed that the ST method produced better recall in the L-H and H-L conditions while the ANT procedure resulted in superior recall in H-H and L-L conditions. Goss and Nodine (1965; Experiment 8) also reported an interaction of methods and stimulus or response similarity in backward recall. The design of this experiment was identical to their seventh experiment with the exception of the substitution of similarity for the meaningfulness variable. No effect due to MP was significant with the acquisition measures, but superiority of backward recall under the ANT procedure in all similarity conditions except L-H was obtained.

A lack of differential retroactive forgetting with different MP was reported by Underwood, Shaughnessy and Zimmerman (1972). Subjects learned two VD lists. The second list was a reversal function of the first list (the right item of the first list became the wrong item of the second and the wrong items of the first functioned as the right items of the second).

In addition, two list lengths were used under each condition. Three trials on the first list and either three or nine trials on the second list were presented. Following the completion of interpolated learning, if any, subjects were given a recall test similar to a test trial in the ST procedure and 5 relearning trials. Neither the recall measure nor the relearning measures resulted in differential effects of MP, though retroaction inhibition was found to be stronger with nine interpolated trials and with the longer list. It will be remembered that superiority of the ST procedure in acquisition performance was shown with the longer of the two lists.

Reports of comparison of retnetion in the PRO versus ANT procedures are limited to two and differ too much in methodology to allow a meaningful conclusion. Stolurow and Lippert (1964) examined retention of a PA task after 24 hours, 7 or 30 days. Acquisition was presented under PRO and ANT procedures to children serving as subjects. Two levels of acquisition were employed. The retention data after 24 hours revealed superiority of the PRO method with the low level of learning but the ANT method was superior with the high level of learning. Similar results were obtained with the longer retention intervals. Silberman, Melaragno and Coulson (1961) compared immediate recall between a variation of the PRO and ANT procedures of multiple choice PAL. The tasks differed from the conventional ANT and PRO procedures only in that questions were used as stimulus terms and multiple choice answers as responses. Unlike Stolurow and Lippert, these authors found no differences in recall between the two methods.

The literature on retention measures as a function of methods of presentation in resolving the sources of discrepancy in the acquisition literature on methods of presentation. It only adds another areas of confusion to an already puzzling literature.

GENERAL CONCLUSION

The three major hypotheses which have been proposed in an attempt to explain the general superiority of the ST method have all shown substantial deficiencies in dealing with the empirical findings, though partial support of each theory has been reported. It should again be noted, however, that most of the studies reviewed have been evaluated for their theoretical implications via hindsight, or a post hoc analysis, and were not specifically designed as tests of the theoretical formulations of more recent vintage.

The reported interaction of list length with MP and the effect of an item's feedback-anticipation position in the list on performance leaves little doubt that the shorter retention interval is indeed one contributing factor in the facilitative effect of the ST procedure. However, the retention interval hypothesis fails to explain the interaction of many other variables (e.g., list material, transfer paradigms, personality factors) with MP and is clearly too simplistic in nature to handle the complexity of empirical phenomena associated with methods of presentation.

Similarly the recent emphasis of Battig (1973) on the temporal separation of learning and performance as a major variable responsible for the superiority of the ST method seems to be insufficient. Izawa's (1971) demonstration that a longer ITI enhances performance in the ST method while it has no effect under the ANT procedure presents a convincing case for the inclusion of ITI as an influential variable in the outcome of results of MP. But method of presentation phenomena in contrast to the ANT and ST methods cannot be regarded as simply a case of a massed versus a distrubted practice effect.

The most promising hypothesis seems to be the "interference hypothesis" forwarded in the earlier formulations of Battig and his associates and extended by the Kanak research group. The extensive scope of predictions that this hypothesis is able to formulate makes a thorough test of the hypothesis possible. Unfortunately, the empirical findings have failed to present strong support of the hypothesis. The inconsistent results of investigations dealing with the effects of list material, rate of presentation, and transfer paradigms on MP, prevents a conclusive statement of support for the hypothesis. However, the source of inconsistent results with each of these variabiles may well lie in the use of different levels of other variables that affect MP. For example, the lack of a predicted interaction of rate of presentation with ST and ANT MP reported by Cofer et al. (1967) may be due to the large ITI of 30 sec employed. The substantial effect of the ITI on the ST method could have minimized a possible weaker effect of rate of presentation of anticipation and feedback periods. Factorial investigations of variables, predicted to differentially affect MP are few. Such investigations are much needed and could increase our understanding of the mechanisms involved in the two MP to a great degree. Nonetheless, the interference hypothesis cannot account for the interaction of list length and MP and the inclusion of Izawa's interpretation of the retention interval on performance would strengthen the hypothesis.

Although the interference hypothesis stresses the role of competition between storage and retrieval processes as the main factor producing the MP effect, it fails to clearly specify the mechanisms involved in such competition. The limited knowledge of specific mechanisms of storage and retrieval processes, at this stage, makes it difficult to identify the means by which such interference takes place. Hypotheses or explanations currently in vogue with respect to free recall task phenomena have not been successfully extended to the PA task, for example, in sufficient depth to alleviate this problem but concepts such as "output interference," encoding variability, etc. could be fruitfully employed. Nonetheless, many enlightening findings in regard to storage and retrieval mechanisms have yet to be included in discussions of MP.

A number of investigators have demonstrated that forgetting may be caused solely by lack of accessibility, and not lack of availability, of the items to be recalled. The advantage of cued recall, when compared with uncued recall, in free recall learning (FRL) has been generally established (see Birnbaum & Eichner, 1971; Tulving & Pearlstone, 1966; Tulving & Osler, 1968). Tulving and Psotka (1971) have reported retroaction effects in FR learning to be mainly due to inaccessibility of materials since cued recall diminishes the effects of retroactive inhibition. The inferiority of the ANT method may very well be caused by inaccessibility of responses. This notion is strengthened in view of Kanak and Neuner's results showing that the two methods of ANT and ST did not differ in the length of the learning stage but did in the associative stage. This finding indicates that the "interference" between storage and retrieval processes affects the retrieval mechanism more potently than it does the sotrage mechanism. Although PA learning in the ANT and ST methods is by nature a "cued" recall, presentation of additional cues during the feedback and anticipation periods may decrease the inferiority of performance under the ANT method.

Another overlooked area of investigation is the effect of study and test trials on retrieval mechanisms. Tulving (1967) varied the number of test trials per study trials in FRL and concluded that, though the number of items recalled on successive test trials did not differ, test and study trials seemed to have equal effects on performance. Rosner (1970) reported similar results. Bergman and Wiener (1970) also varied the number of test trials per study trial (either alternating study and test trials or one study trial followed by three test trials) in FRL and PAL and found an interaction of task and number of test trials. Although increasing the number of test trials improved performance in FRL, it had no effect on PA performance. However, extensive investigation of the effects of test and study trials on PA performance (Izawa, 1966; 1967; 1968; 1969) indicates that test trials do aid performance in PAL and cannot be regarded as neutral trials. For example, Izawa (1968) found performance was better when each study trial was followed by two test trials than when study and test trials alternated. This investigation pointed out that the difference between test and study trials are qualitative rather than quantitative. That is, although "learning, in the sense of a systematic increase in correct response probability, does not occur in nonreinforced test trials" (Izawa, 1966; p. 917), the function of test trials remains to be that of potentiation of subsequent study trials (Izawa, 1968), Thus, the efficiency of storage mechanims seems to increase with the increase in the number of interspersed test trials. The length of "test trials" in the ST method is substantially greater than that of the ANT method which is limited to one. It is, therefore, to be expected that the potentiating effect of test trials on storage processes in the ST method is to be lacking in the ANT procedure. This lack may be a factor leading to the inferiority of the ANT procedure. If so, increasing the number of anticipation periods in relation to feedback periods should lead to improvement in performance in the ANT method and should reduce its inferiority to the traditional ST procedure.

Psychologists have traditionally shown a "significant" lack of interest in methodological variables and their potential for interaction with "processes." Admittedly, investigations of methodological factors per se provide less satisfaction than examination of more closely related behavioral factors. Tulving and Madigan (1970) have recently separated "students of verbal learning" from "students of memory" and have criticized the former for their obsession with tasks for tasks' sake. Obviously a preoccupation with tasks per se may result in the neglect of the true purpose of verbal <u>learning</u> research. However, the foundation of every science is based on sound methodology and a knowledge of how methodology interacts with phenomena. Strong methodological fcundations are still lacking in many areas of psychology, including the psychology of learning. Investigations of methodological effects, albeit not as exciting, are as necessary as inquiries into behavioral factors. Such research may provide a step toward unification of students of "verbal learning" and "memory" rather than their separation.

REFERENCES

Aiken, E. G. and Lau, A. W. Response prompting and response confirmation: A review of recent literature. Psychological Bulletin, 1967, 68, 330-341.

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs, 1960, 74 (10 whole No. 497).

Battig, W. F. Procedural problems in paired-associate learning research. Psychonomic Monograph Supplement, 1965, No. 1.

Battig, W. F. Advantage of recall over anticipation method in verbal learning. Behavior Research Methods and Instrumentation, 1969, 1, 217-222.

Battig, W. F. Evidence that recall-anticipation differences increase with increasing time interval. Journal of Experimental Psychology, 1973, 1973, 100, 393-402.

Battig, W. F. And Brackett, H. R. Comparisons of anticipation and recall method of paired-associate learning. Psychological Reports, 1961, 9, 59-65.

Battig, W. F. and Brackett, H. R. Transfer from verbal-discrimination to paired-associate learning: II. Effects of intralist similarity, method and percentage occurrence of response members. Journal of Experimental Psychology, 1963, 65, 507-514.

Battig, W. F., Brown, S. C. and Nelson, D. Constant vs. varied serial order in paired-associate learning. Psychological Reports, 1963, 12, 695-721.

Battig, W. F. and Lawrence, P.S. The greater sensibility of the serial recall than anticipation procedure to variations in serial order. Journal of Experimental Psychology, 1967, 73, 172-178.

Battig, W. F. and Switalski, R.W. Comparison of anticipation and recall procedures in verbal-discrimination learning. Psychonomic Science, 1966, 6, 65-66.

Battig, W. F., Williams, J.M. and Williams, J.G. Transfer from verbal-discrimination to paired-associate learning. Journal of Experimental Psychology, 1962, 63, 258-268.

Battig, W.F. and Wu, R.D. Comparison of recall and anticipation paired-associate procedure with mixed orally-presented lists. Psychonomic Science, 1965, 3, 233-234.

Behring, O.W. and Zaffy, D.J. Study-test and anticipation methods in paired-associate learning of similar and nonsimilar lists. Psychological Reports, 1965, 16, 1237-1241.

Bergman, A. S. and Wiener, J. R. Effects of test trials in paired-associate and free recall learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 689-698.

Birnbaum, I. M. and Eichner, J.T. Study versus test trial and long-term retention in free recall learning. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 516-521.

Blue, M.C. Anticipation vs. recall in paired-associate learning of intellectually average and retarded subjects. Psychological Reports, 1970, 27, 327-334.

Cavanagh, J.P. and Parkman, J.A. Comparison of the study-recall and anticipation method in steady state paired-associate learning. Psychonomic Science, 1971, 22, 361-363.

Cattell, R. B. and Eber, H. W. The sixteen personality factor questionnaire. Illinois: Institute for Personality and Ability Testing, 1962.

Champion, R. A. and McBride, D. A. Activity during delay of reinforcement in human learning. Journal of Experimental Psychology, 1962, 63, 589-592.

Cofer, N.C.; Diamond, F.; Olson, R.A.; Stein, J.S.; and Walker, M. Comparison of anticipation and recall methods in paired-associate learning. Journal of Experimental Psychology, 1967, 75, 545-558.

Cook, J.O. Supplementary report: Processes underlying learning a single paired-associate item. Journal of Experimental Psychology, 1958, 56, 455.

Cook, J.O. and Kendler, T.S. A theoretical model to explain some paired-associate learning data. In G. Finch and F. Cameron (eds.), Symposiums on Air Force human engineering personnel and training research. Washington, D.C.: National Academy of Science--Nation Research Council, 1956, pp. 90-98.

Cook, J.O. and Spitzer, M.E. Supplementary report: Prompting versus confirmation in paired-associate learning. Journal of Experimental Psychology, 1960, 59, 275-276.

Duffy, J.C. and Kanak, N.J. Confidence thresholds in two methods of paired-associate learning as a function of ego-strength. American Journal of Psychology, in press.

Eckert, E. and Kanak, N.J. Verbal discrimination learning: A review of the acquisition, transfer, and retention literature through 1972. Psychological Bulletin, 1974, 81, 282-607.

Elmes, D.G. and Lovelace, E.A. Meaningfulness and presentation method in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 296-297.

Ekstrand, B.R. A note on measuring response learning during paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 344-347.

Ekstrand, B.R., Wallace, W.P. and Underwood, B.J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 566-578.

Fulkerson, F.E. and Johnson, J.E. Methodological variables in verbal-discrimination learning. Psychonomic Science, 1971, 22, 68-69.

Gagne, R.M. The effect of sequence of presentation of similar items on learning of paired-associates. Journal of Experimental Psychology, 1950, 40, 61-73.

Gasparikova, M. Paired-associate acquisition of competing and noncompeting material as a function of anticipation versus study-test methods of presentation. Unpublished doctoral dissertation, University of Oklahoma, 1972.

Glaze, J.A. The association value of nonsense syllables. Journal of Genetic Psychology, 1928, 35, 255-267.

Gorman, A.M. Recognition of memory for nouns as a function of abstractness and frequency. Journal of Experimental Psychology, 1961, 61, 23-29.

Goss, A.E. and Nodine, C.F. Paired-associate learning. New York: Academic Press, 1965.

Hawker, J.R. The influence of training procedure and other task variables in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1964(a), 3, 70-76.

Hawker, J.R. The effect of training procedure, response availability and response meaningfulness in multiple choice paired-associate learning. Psychonomic Science, 1965(a), 3, 329-330.

Hawker, J.R. The effects of training procedure, response similarity, and number of response after natives in multiple-choice paired-associate learning. Psychonomic Science, 1965(b), 3, 331-332.

Hawker, J.R. The effect of training procedure and variations in paired-associate paradigm on the learning of paried-associates. Psychonomic Science, 1967, 7, 347-348.

Hilgard, E.R. Methods and procedure in the study of learning. In S.S. Stevens (ed.), Handbook of experimental psychology. New York: Wiley, 1951, pp. 517-567.

Horton, D.L. and Wiley, R.E. Mediated facilitation comparison of two methods and two paradigms. Psychological Reports, 1967, 21, 3-7.

Ingison, L.J. and Ekstrand, B.R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. Journal of Experimental Psychology, 1970, 85, 249-254.

Izawa, C. Reinforcement test sequence in paired-associate learning. Psychological Reports, 1966, 18, 879-919.

Izawa, C. Function of test trials in paired-associate learning. Journal of Experimental Psychology, 1967, 75, 194-209.

Izawa, C. Effects of reinforcement method and test trials upon paired-associate acquisition and retention. Psychological Reports, 1968, 23, 947-959.

Izawa, C. Comparison of reinforcement and list trials in paired-associate learning. Journal of Experimental Psychology, 1969, 81, 600-603.

Izawa, C. List versus items in distribution practice in paired-associate learning. Proceedings of the Annual Convention of the American

Psychological Association, 1970, 5, 87-88.

Izawa, C. Massed and spaced practice in paired-associate learning: List versus item distributions. Journal of Experimental Psychology, 1971, 89, 10-21.

Izawa, C. Retention interval hypothesis and evidence for its basic assumptions. Journal of Experimental Psychology, 1972, 96, 17-24.

Izawa, C. Retention interval hypothesis and list lengths: Comparison of anticipation and study-test procedures. Canadian Journal of Psychology, 1974, 28, 204-224.

Jones, R.E. Effects of delay of informative feedback, post-feedback interval and feedback presentation mode on verbal paired-associate learning. Journal of Experimental Psychology, 1968, 77, 87-93.

Jones, R.E. and Bourne, L.E. Delay of informative feedback in verbal learning. Canadian Journal of Psychology, 1964, 18, 266-280.

Kanak, N.J., Cole, L.E. and Eckert, E. Implicit associative responses in verbal discrimination acquisition. Journal of Experimental Psychology, 1972, 93, 309-319.

Kanak, N.J. and Neuner, S.D. Associative symmetry and item availability as a function of five methods of paired-associate acquisition. Journal of Experimental Psychology, 1970, 86, 288-295.

Kausler, D.W. Comparison of anticipation and recall methods for geriatric subjects. Psychological Reports, 1963, 13, 702.

Kausler, D.W. Psychology of verbal learning and memory. New York: Academic Press, 1974.

Kopstein, F.F. and Roshal, S.M. Verbal learning efficiency as influenced by manipulation of presentational response processes: Picturial-verbal and temporal contiguity factors. In A.A. Lumsdaine (ed.), Student response in programmed instruction. Washington, D.C.: National Academy of Sciences-Natural Research Council, 1961, pp. 335-350.

Kopstein, F.F. and Roshal, S.M. Method of presentation word pairs as a factor in foreign vocabulary learning. American Psychologist, 1955, 10, 354.

Levine, J.M. Prompting and confirmation as a function of familiarity of stimulus materials. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 421-424.

Lintz, L.M. and Brackbilly, Y. Effects of reinforcement delay during learning on retention of verbal material in adults, Journal of Experimental Psychology, 1966, 71, 194-199.

Lockhead, G.R. Method of presenting paired associates. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 62-65.

Martin, E. and Melton, A.W. Meaningfulness and trigram recognition. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 126-135.

Moursund, J.P. and Chape, C.M. Comparison of three methods in paired-associate learning. Psychonomic Science, 1966, 6, 167-168.

Noble, C.E. Measurements of association value (a) rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of English alphabet. Psychological Reports, 1961, 8, 487-521.

Nodine, C.F. and Nodine, B.F. Stimulus intervals, stimulus durations and difficulty level in paired-associate learning. Journal of Experimental Psychology, 1966, 72, 156-158.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A., Yuille, J.C. and Madigan, S.A. Concreteness, imagery and meaningfulness values of 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Peacock, L.G. Paired associate transfer as a function of anticipation versus free learning. Unpublished doctoral dissertation, University of Oklahoma, 1971.

Pennington, L.A. and Walter, R.H. The anticipation method in paired associate learning. Journal of Psychology, 1938, 6, 281-283.

Peterson, L.R. and Brewer, C.I. Confirmation, correction and contingity. Journal of Verbal Learning and Verbal Behavior, 1963, i, 365-371.

Posnansky, C.J. Probing for the functional stimuli in serial learning. Journal of Experimental Psychology, 1972, 96, 184-193.

Reynolds, J.H. Effects of non-anticipation and anticipation procedures upon paired-associate learning in mixed and unmixed list design. Psychological Reports, 1964, 15, 795-801.

Reynolds, J.H. Confirmation, contingity and response practice in paired-associate learning. Journal of Experimental Psychology, 1967, 73, 394-400.

Rosner, S.R. The effect of presentation and recall trials on organization in multitrial free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 69-74.

Rothkopf, E.Z. and Coke, E.V. Repetition interval and rehearsal method in learning equivalences from written sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 406-416.

Rowe, E.J. and Paivio, A. Effects of noun imagery, pronounciation, method of presentation and intrapair order of items in verbal discrimination. Journal of Experimental Psychology, 1972, 93, 427-429.

Sampson, J.F. Delay of reinforcement in human verbal learning. Australian

Journal of Psychology, 1971, 23, 35-45.

Schild, M.E. and Battig, W.F. Directionality in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 42-49.

Sidowski, P.B., Kopstein, F.F. and Shillestad, I.J. Prompting and confirmation variables in verbal learning. Psychological Reports, 1961, 8, 401-406.

Silberman, H.F., Melaragno, R.J. and Coulson, J.E. Confirmating and prompting with connected discourse material. Psychological Reports, 1961, 9, 235-238.

Spence, K.W. Behavior theory and conditioning. New Haven: Yale University Press, 1956.

Spence, K.W., Farber, I.T. and McFann, H.H. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. Journal of Experimental Psychology, 1956, 52, 296-305.

Stolvrow, L.M. and Lippert, H.T. Prompting, confirmation, and overlearning in automated teaching of sight vocabulary. In J.P. DeCecco (ed.) Educational technology. New York: Holt, Rinehart & Winston, 1964, pp. 187-197.

Tulving, E. The effects of presentation and recall of material in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1967, 7, 175-184.

Tulving, E. and Arbuckle, T.Y. Sources of intratrial interference in Immediate recall of paired associates. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 321-334.

Tulving, E. and Madigan, S.A. Memory and verbal learning. In P.H. Mussen and M.R. Rosenzweig (eds.), Annual review of psychology, Vol. 21. Palo Alto, California: Annual Reviews, 1970.

Tulving, E. and Osler, S. Effectiveness of retrieval cues in memory for words. Journal of Experimental Psychology, 1968, 77, 593-601.

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

Tulving, E. and Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. Journal of Experimental Psychology, 1971, 87, 1-8.

Underwood, B.J. Ten years of massed practice on distributed practice. Psychological Review, 1961, 68, 229-247.

Underwood, B.J. and Ekstrand, B.R. Effect of distributed practice on paired-associate learning. Journal of Experimental Psychology, 1967, 73, (4, pt. 2).

Underwood, B.J. and Goad, D. Studies of distributed practice: I. The influence of intralist similarity in serial learning. Journal of Experimental Psychology, 1951, 42, 125-134.

Underwood, B.J., Runquist, W.N. and Schilz, R.W. Response learning in paired-associate lists as a function of intralist similarity. Journal of Experimental Psychology, 1959, 58, 70-78.

Underwood, B.J. and Schulz, R.W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.

Underwood, B.J., Shaughnessy, J.J. and Zimmerman, J. List length and method of presentation in verbal discrimination learning with further evidence on retraction. Journal of Experimental Psychology, 1972, 93, 181-187.

Voss, J.F. Recall and anticipation methods in probabilistic associative learning. Behavior Research Methods and Instrumentation, 1967, 1, 111-114.

Ward, L.C. and Maisto, A.A. The effect of delay, and direction of feedback on verbal-discrimination learning. American Journal of Psychology, 1973, 86, 547-554.

Wright, J.H. Effect of stimulus meaningfulness, method of presentation and list design on the learning of paired associates. Journal of Experimental Psychology, 1967, 73, 72-77.

TABLE 1

Range and mean of each of the four parameters identified by Izawa (1972).

Type of parameter	R-R		T-T		T-T		R-T	
MP	ANT	ST	ANT	ST	ANT	ST	ANT	ST
Range	1,3n-3	n,3n-2	1,4n-3	n,3n-2	0	0,2n-2	0,4n-4	0,2n-2
mean	2n-1	2n-1	2n-1	2n-1	0	n-1	2n-2	n-1

TABLE 2

Mean number of incorrect responses on each of the 3 test trials reported by Sidowski _et_ _al_ (1961).

Group	Tests		
	1	2	3
PRO	8.65	5.00	5.00
PRO-SR	6.05	2.90	3.85
ANT	10.70	7.70	7.05
ANT-SR	10.15	6.55	5.30
ANT-PRO	9.35	6.35	5.50
SIM	6.90	4.00	3.50

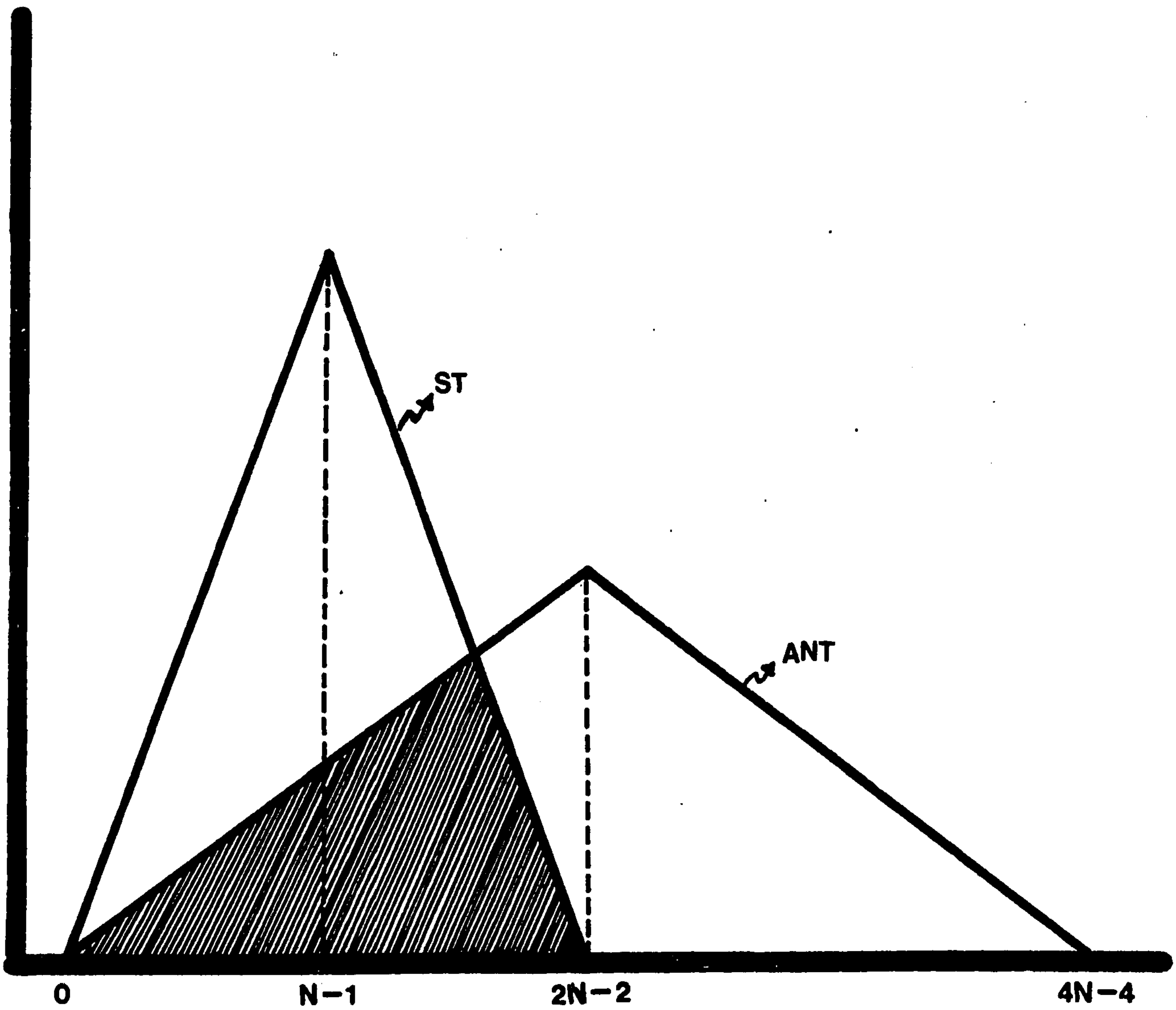
TABLE 3

The designs of the experiments reported by Izawa (1971)

	Experiment number								
	I	II	III	IV	V	VI	VII	VIII	IX
Types of interval	ITI	I-item-I	ITI	I-item-I	ITI	I-item-I	ITI	I-item-I	I-item-I
Locus of interval	---	---	study	study	test	test	both	both	both
MP	ANT	ANT	ST	ST	ST	ST	ST	ST	ST
Types of interval filler	N	N	N	N	N	N	N	N	B

FIGURE 1

Distribution curves of n R-T intervals as a function of intervening events under a random item presentation (adapted from Izawa, 1972).


ST
ANT
0
N−1
2N−2
4N−4

APPENDIX B

LIST

Two Types of Lists Presented to the Subject

1		2	
BUX	26	BUV	28
CAX	51	CAJ	54
DIJ	31	DIH	37
FEX	35	FEH	32
GEK	49	GEC	48
HUV	93	HUY	91
JAH	38	JAF	39
KIQ	84	KIB	86
LUY	24	LIY	27
MOJ	56	MOQ	52
NAF	96	NAX	95
NUX	63	NUH	62
PIB	71	PUG	74
QOR	87	QON	83
QUZ	64	QUR	65
RIH	23	RIJ	29
SOQ	67	SOJ	61
TEH	92	TEF	94
VAJ	59	VAY	53
VON	43	VOP	41
WEF	47	WEZ	46
YIL	79	YIT	78
YOD	85	YOC	89
ZAS	72	ZAK	76

APPENDIX C

INSTRUCTIONS

INSTRUCTIONS

Instructions for Subjects in IF-1 Condition.

This experiment deals with learning of verbal materials. You will be presented a list containing 24 pairs. Each pair consists of a three-letter nonsense syllable and a two-digit number. In general you see the items presented either as a syllable-number pair or as a syllable presented alone. When a syllable and its paired number appear together, you should just study them closely but quietly. WHEN A SYLLABLE APPEARS ALONE YOU SHOULD TRY TO REMEMBER THE NUMBER WHICH IS PAIRED WITH IT AND SAY THE NUMBER ALOUD.

In the first trial of the experiment all the 24 pairs will be presented, one at a time, for two seconds each pair. Study the pairs quietly. For the remaining trials in the experiment, the items are presented in the following manner. First a nonsense syllable will appear alone for 2 seconds. During this period you should try to remember the number which is paired with the nonsense syllable and say the number loudly so I can hear it. Right after that the same nonsense syllable will appear along with its paired number, again for two seconds. This gives you a chance to study the pair again. Following this the next nonsense syllable will be presented first alone for 2 seconds and then with its paired number for an additional 2 seconds.

This procedure continues until all 24 pairs in the list are presented. The list will be presented again and again. You can make a guess when you're not sure of the correct number, but be sure to give the number in the 2-second period when the nonsense syllable is presented alone and before the pair appears. The list starts with a series of stars and ends with a blank space.

Remember to respond with a number whenever a nonsense syllable is presented alone and make no response whenever a nonsense syllable and a number are presented together. Below is an example of how the items will appear.[1]

Instructions for Subject in DF-1 Condition.

This experiment deals with learning of verbal materials. You will be presented a list containing 24 pairs. Each pair consists of a three-letter nonsense syllable and a two-digit number. In general you see the items presented either as a syllable-number pair or a syllable presented alone. When a syllable and its paired number appear together, you should just study them closely but quietly. WHEN A SYLLABLE APPEARS ALONE YOU SHOULD TRY TO REMEMBER THE NUMBER WHICH IS PAIRED WITH IT AND SAY THE NUMBER ALOUD.

In the first trial of the experiment all the 24 pairs will be presented, one at a time, for two seconds each pair. Study the pairs quietly. For the remaining trials in the experiment, the items are presented in the following manner. First a nonsense syllable will appear alone for 2 seconds. During this period you should try to remember the number which is paired with the nonsense syllable and say the number loudly so I can hear it. Right after that a different pair consisting of a syllable and a number will appear, again for two seconds. This gives you a chance to study that pair. Following this first the next nonsense syllable will be presented alone for 2 seconds and then another pair will appear for an additional 2 seconds.

This procedure continues until all 24 pairs in the list are presented. The list will be presented again and again. You can make a guess when you're not sure of the correct number, but be sure to give the number in the 2-second period when the nonsense syllable is presented alone and before a pair appears. The list starts with a series of stars and ends with a blank space.

Remember to respond with a number whenever a nonsense syllable is presented alone and make no response whenever a nonsense syllable and a number are presented together. Below is an example of how the items will appear.[1]

Instructions for Subjects in Conditions IF-2 -12.

This experiment deals with learning of verbal materials. You will be presented a list containing 24 pairs. Each pair consists of a three-letter nonsesnse syllable and a two-digit number. In general you see the items presented either as a syllable-number pair or as a syllable presented alone. When a syllable and its paired number appear together, you should just study them closely but quietly. WHEN A SYLLABLE APPEARS ALONE YOU SHOULD TRY TO REMEMBER THE NUMBER WHICH IS PAIRED WITH IT AND SAY THE NUMBER ALOUD.

In the first trial of the experiment all the 24 pairs will be presented, one at a time, for two seconds each pair. Study the pairs quietly. For the remaining trials in the experiment, the items are presented in the followig manner. First * nonsense syllables will be presented, one at a time, for a period of 2 seconds each. During each of these 2-second periods you should try to remember the number which is paired with the paritcular nonsense syllable you see and say the number loudly so I can hear it. After the * nonsense syllables are presented, the same * nonsense syllables will appear along with their paired number. The pairs will appear one at a time for 2 seconds per pair, but not necessarily in the same exact order that their corresponding nonsense syllables had appeared. Following this the next set * nonsense syllables will be presented first alone for 2 seconds each and then with their corresponding numbers again for 2 seconds per pair.

This procedure continues until all 24 pairs in the list are presented. The list will be presented again and again. You can make a guess when you're unsure of the correct number, but be sure to give the number in the 2-second period when the nonsense syllable is presented alone and before the next item appears. The list starts with a series of stars and ends with a blank space.

Remember to respond with a number whenever a nonsense syllable is presented alone and make no response whenever a nonsense syllable and a number are presented together. Below is an example of how the items will appear.[1]

Instructions for Subjects in Conditions DF-2 -12.

This experiment deals with learning of verbal materials. You will be presented a list containing 24 pairs. Each pair consists of a three-letter nonsense syllable and a two-digit number. In general you see the items presented either as a syllable-number pair or as a syllable presented alone. When a syllable and its paired number appear together, you should just study them closely but quietly. WHEN A SYLLABLE APPEARS ALONE YOU SHOULD TRY TO REMEMBER THE NUMBER WHICH IS PAIRED WITH IT AND SAY THE NUMBER ALOUD.

In the first trial of the experiment all the 24 pairs will be presented, one at a time, for two seconds each pair. Study the pairs quietly. For the remaining trials in the experiment, the items are presented in the following manner. First * nonsense syllables will be presented, one at a time, for a period of two seconds each. During each of these 2-second periods you should try to remember the number which is paired with the particular nonsense syllable you see and say the number loudly so I can hear it. After the * nonsense syllables are presented, * different syllable-number pairs will appear. The pairs will appear one at a time for 2 seconds per pair. This gives a chance to study these pairs. Following this, first the next set of * nonsense syllables will be presented alone each for a period of 2 seconds and then * different pairs will appear at the same rate.

This procedure continues until all 24 pairs in the list are presented. The list will be presented again and again. You can make a guess when you're not sure of the correct number, but be sure to give the number in the 2-second period when the nonsense syllable is presented alone and before the next item appears. The list starts with a series of stars and ends with a blank space.

Remember to respond with a number whenever a nonsense syllable is presented alone and make no response whenever a nonsense syllable and a number are presented together. Below is an example of how the item will appear.[1]

Instructions for Subjects in Conditions IF-24 and DF-24.

This experiment deals with learning of verbal materials. You will be presented a list containing 24 pairs. Each pair consits of a three-letter nonsense syllable and a two-digit number. In general you see the items presented either as a syllable-number pair or as a syllable presented alone. When the syllable and its paired number appear together, you should just study them quietly. WHEN A SYLLABLE APPEARS ALONE, YOU SHOULD TRY TO REMEMBER THE NUMBER WHICH IS PAIRED WITH IT AND SAY THE NUMBER ALOUD.

The items in the list will be presented in the following manner. First all the 24 pairs will appear, one at a time, for 2 seconds per pair. Study the pairs quietly. After all the pairs are presented, all the 24 syllables will be presented again one at a time at the rate of two seconds per item. During each of these 2-second periods you should try to remember the number which is paired with the particular nonsense syllable you see and say the number loudly so I can hear it. Following the presentation of all 24 syllable alone, all the pairs will be presented again in the same manner. This alternating sequence will appear again and again. Each sequence starts and ends with a series of stars.

You can make a guess when you're not sure of the correct number, but be sure to give the number in the 2-second period when the nonsense syllable is presented alone and before the next item appears.

Remember to respond with a number whenever a nonsense syllable is presented alone and make no response whenever a nonsense syllable and a number are presented together. Below is an example of how the items will appear.[1]

---

[1]Appropriate examples for each condition were then shown to the suject.

*The number of items per grouping for each condition was written in the blank space.

APPENDIX D

SUMMARIES OF ANALYSIS OF VARIANCE

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Total Number of Correct

Responses Per Block of Four Trials

Source	MS	df	F	P
Total	613.350	959		
Between	1354.799	159		
A (Feedback)	3131.891	1	2.5295	.1099
B (Method of Presentation)	4329.574	7	3.4969	.0020
AB	526.163	7	.4250	.8857
%E (error)	1238.131	144		
Within	465.987	800		
C (Block)	60255.047	5	709.4290	.0000
AC	261.875	5	3.0833	.0094
BC	177.375	35	2.0884	.0005
ABC	81.446	35	.09589	.5381
%E (error)	84.935	720		

Summary of (Feedback x Method of Presentation x Blcok)

Analysis of Variance on Total Number of Intrusion

Errors per Block of Four Trials

Source	MS	df	F	P
Total	239.146	959		
Between	1265.368	159		
A (Feedback)	1865.205	1	1.5217	.2169
B (Method of Presentation)	1833.893	7	1.4962	.1724
AB	1426.977	7	1.1642	.3265
%E (Error)	1225.709	144		
Within	35.184	800		
C (Block)	368.822	5	11.5079	.0000
AC	25.492	5	.7954	.5549
BC	60.623	35	1.8915	.0019
ABC	27.961	35	.8724	.6817
%E (Error)	32.049	720		

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Total Number

of Ommission Errors per Block of Four Trials

Source	MS	df	F	P
Total	691.585	959		
Between	1962.107	159		
A (Feedback)	10018.176	1	5.5968	.0183
B (Method of Presentation)	5160.473	7	2.8830	.0077
AB	1153.935	7	.6447	.7202
%E (Error)	1789.971	144		
Within	439.069	800		
C (Block)	56144.535	5	629.8162	.0000
AC	215.487	5	2.4173	.0341
BC	105.250	35	1.1807	.2209
ABC	45.054	35	.5054	.9925
%E (Error)	89.144	720		

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Total Number of Errors

per Block of Four Trials

Source	MS	df	F	P
Total	612.498	959		
Between	1410.710	159		
A (Feedback)	3139.391	1	2.4146	.1184
B (Method of Presentation)	4328.770	7	3.3293	.0029
AB	519.288	7	.3994	.9014
%E (Error)	1300.188	144		
Within	453.854	800		
C (Block)	61132.063	5	914.6594	.0000
AC	308.612	5	4.6175	.0006
BC	163.982	35	2.4535	.0000
ABC	56.250	35	.8416	.7299
%E (Error)	66.836	720		

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Proportions of

Errors per Block of Four Trials

Source	MS	df	F	P
Total	.065	959		
Between	.141	159		
A (Feedback)	.303	1	2.3554	.1230
B (Method of Presentation)	.453	7	3.5233	.0019
AB	.049	7	.3831	.9110
%E (Error)	.129	144		
Within	.050	800		
C (Block)	6.399	5	669.7290	.0000
AC	.024	5	2.5229	.0278
BC	.020	35	2.0491	.0006
ABC	.008	35	.8395	.7330
%E (Error)	.010	720		

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Proportion of Intrusion

per Block of Four Trials

Source	MS	df	F	P
Total	.054	959		
Between	.251	159		
A (Feedback)	.960	1	3.9219	.0467
B (Method of Presentation)	.313	7	1.2810	.2629
AB	.226	7	.9220	.5076
%E (Error)	.245	144		
Within	.014	800		
C (Block)	.463	5	40.4624	.0000
AC	.014	5	1.2399	.2877
BC	.018	35	1.5967	.0167
ABC	.007	35	.6242	.9573
%E (Error)	.011	720		

Summary of (Feedback x Method of Presentation x Block)

Analysis of Variance on Proportion of

Omissions per Block of Four Trials

Source	MS	df	F	P
Total	.065	959		
Between	.281	159		
A (Feedback)	1.533	1	5.7013	.0173
B (Method of Presentation)	.482	7	1.7929	.0924
AB	.154	7	.5731	.7784
%E (Error)	.269	144		
Within	.022	800		
C (Block)	.962	5	61.4786	.0000
AC	.039	5	2.4670	.0310
BC	.014	35	.9160	.6100
ABC	.014	35	.9002	.6364
%E (Error)	.016	720		

Summary of (Feedback x Method of Presentation x List Form)

Analysis of Variance on Total Number of Errors

Source	MS	df	F	P
Total	8479.891	139		
Between	11738.070	31		
A (Feedback)	19145.012	1	2.4893	.1131
B (Method of Presentation)	26286.445	7	3.4179	.0025
C (List Form)	120.000	1	.0156	.8963
AB	3194.991	7	.4154	.8914
AC	110.000	1	.0143	.9008
BC	13965.723	7	1.8159	.1890
ABC	5787.863	7	.7500	.6318
%E (Error)	7690.805	128		

Summary of (Feedback x Method of Presentation)

Analysis of Variance on Number of Trials

to Reach the Criterion of 8/24

Correct Responses

Source	MS	df	F	P
Total	30.578	159		
Between	49.541	15		
A (Feedback)	25.601	1	.8950	.6522
B (Method of Presentation)	86.087	7	3.0097	.0058
AB	16.415	7	.5739	.7778
%E (Error)	28.603	144		

Summary of (Feedback x Method of Presentation)

Analysis of Variance on Number of Trials

to Reach the Criterion of 4/24

Correct Responses

Source	MS	df	F	P
Total	10.921	139		
Between	16.920	15		
A (Feedback)	.305	1	.0296	.8579
B (Method of Presentation)	27.735	7	2.6937	.0119
AB	8.477	7	.8233	.5706
%E (Error)	10.297	144		

Summary of (Feedback x Method of Presentation)

Analysis of Variance on Number

of Correct Responses on the Last Trial

Source	MS	df	F	P
Total	29.263	159		
Between	58.510	15		
A (Feedback)	100.803	1	3.8451	.0488
B (Method of Presentation)	91.241	7	3.4804	.0021
AB	19.737	7	.7529	.6294
%E (Error)	26.216	144		