The effect of verbal organizations on the location of subjects' errors in serial learning¹

JOEL R. LEVIN, *University of Wisconsin, Madison, Wis. 53706*

The implementation of sentences to organize a 12-item serial list was found to have an effect on Ss' resulting error profiles. Under two methods of testing, namely serial anticipation and serial recall, profiles based respectively on serial-position errors and transitional errors were found to differ from those obtained when verbal organizations were not employed.

The serial-position effect is certainly among the most characteristic of human learning phenomena. Its presence has been revealed consistently in serial learning tasks involving sequential associations among stimulus members. Too, the phenomenon is known to persist over a wide variety of S types, learning materials, tasks, and instructional sets (see Mednick, 1964).

There is reason to believe, however, that the familiar bow-shaped curve is more typical of simple rote-learning situations than those in which materials are presented in a meaningful, well organized manner. Examples of differently organized materials may be found in the classic Miller & Selfridge (1950) study in which, in one experimental condition, words were randomly selected from all grammatical form classes and subsequently strung together in serial fashion, while in others, words were selected and placed together sequentially under conditions of varying approximations to English grammaticality.

Approximating Miller and Selfridge's experimental conditions, Simpson (1965, 1967) has found that verbal sequences which are more sentence-like in nature are better recalled (in serial order) than those with lesser resemblance to sentence word-order. Furthermore, Simpson has noted that the serial-position curves obtained under the two kinds of presentations are, to some extent, different. One might hypothesize that better organized learning materials like, for example, meaningful English sentences, will produce deviations from the anticipated smooth bow-shaped curve typically obtained under pure rote-learning conditions.

The preceding hypothesis has been tested in the present experiment, in which all to-be-learned items were exclusively nouns, with organizations of the list varying according to the type of accompanying

verbalization presented with the printed nouns. That is, identical lists of nouns were presented in the context of one or more sentences, with the question of interest focussing on the relationship between type of organization provided and error profile obtained, according to item location.

METHOD²

A list of 12 high-frequency nouns was presented to Ss under one of seven experimental conditions. In the sentence conditions, all or portions of the list were organized via provided verbal material. Further distinctions among these conditions were made by the number of discrete sentences incorporated as organizers: one (S1), two (S2), three (S3), four (S4), six (S6), and 12 (S12). In the S1 condition, for example, a single sentence was used to tie together all of the items in the list, while in S2, two unrelated sentences were employed (one tying together the list's first six items, and one, the last six items). Alternately, in the S12 condition, 12 unrelated sentences provided verbal contexts for each of the 12 nouns, but in no way tied them together with other items in the list. The final experimental condition consisted of a control group (S0), in which the 12 nouns were presented in the absence of accompanying verbal material.

All Ss were selected from the population of fourth- and fifth-grade children at an elementary school serving an upper-middle-class residential area, and were randomly assigned to one of the experimental conditions. A total of 112 Ss was used, being equally distributed among the seven conditions and further broken down according to method of testing, list, and order.

Method of testing was represented by the anticipation and recall methods, one of which was used to assess a given S's serial learning performance. Additionally, a particular S received one of two 12-item serial lists, presented in one of two sequential arrangements (orders). The two latter variables may be viewed more as external validity controls than as critical components of the present design.

All materials were typed onto tapes which were used in conjunction with a Lafayette memory drum. Each noun to be learned was exposed in the window of the memory drum for a total of 4 sec. On the first study trial, the E pronounced each noun aloud as it appeared, as well as the verbal material

appropriate for the particular experimental condition. Under the anticipation method there was an 8-sec intertrial interval, while under the recall method there were two 4-sec intervals, one between each study and test trial and one between each test and study trial. Following the initial study trial, all nouns were presented without companion verbal material (as in S0), and E ceased naming the items as they appeared.

All Ss were run to a criterion of an errorless recitation of the list; however, no S was given more than 15 test trials. Among the dependent measures taken for each S was number of errors (to criterion), according to their location within the list.

RESULTS AND DISCUSSION

Of the 112 participating Ss, four failed to reach criterion. Of these, one apiece came from the S4, S6, S12, and S0 conditions. For these Ss, errors by item location were based on their performance through 15 test trials.

The results will be presented according to method of testing. (As will be seen presently, an index other than serial-position errors was adopted to describe Ss' error profiles under the recall method.)

Anticipation

In order to make comparisons of serial-position error profiles among experimental conditions, the number of errors made by each S at the 12 item positions was determined. Then the item positions were rank-ordered according to the error frequency at each, resulting in a unique position rank-ordering for each S ranging from 1 (lowest error occurrence) through 12 (highest error occurrence).³ In cases where the same number of errors were made at two or more positions, the method of midranks (averaging the ranks over which the ties occurred and assigning this average to each of the tied positions) was adopted.

The average ranks of errors at each position, according to experimental condition, may be found in Table 1. It should be remembered that low-average ranks correspond to few errors, while high-average ranks correspond to many errors. A glance at the data in Table 1 reveals that for most conditions, the average ranks do not appear equal across item positions. The degree to which this is true has been determined statistically, for each experimental condition, by means of the nonparametric Friedman test for correlated observations (Friedman, 1937).

Moreover, in the present experiment it was deemed important to determine the properties of each serial-position curve obtained under the various experimental conditions. Marascuilo & McSweeney (1967) have recommended the use of orthogonal polynomials in tests for trend with Friedman-type data. Therefore, the

Table 1
Average Ranks of the Errors at the 12 Item Positions, According to Experimental Condition (Anticipation Method, Eight Ss per Condition)

Condition	Item Position Number											
	1	2	3	4	5	6	7	8	9	10	11	12
S1	2.50	5.38	8.69	7.00	10.06	8.00	5.06	8.06	6.69	7.12	5.31	4.12
S2	2.00	6.00	5.69	8.25	8.12	9.00	7.06	7.25	6.94	7.44	5.88	3.69
S3	2.12	4.94	6.06	6.62	7.75	8.62	9.44	7.19	7.50	8.00	5.44	4.31
S4	3.56	6.31	6.88	7.25	6.00	7.50	7.25	7.12	6.56	8.12	6.25	5.19
S6	2.44	4.50	5.50	7.88	8.00	8.06	9.19	9.06	8.69	5.00	5.88	3.81
S12	1.69	4.88	6.06	5.81	10.19	9.44	8.44	6.81	7.12	5.94	6.19	5.44
S0	1.81	2.75	4.88	5.75	8.88	9.62	7.19	8.19	8.25	7.31	7.88	5.50

Table 2
Chi-Squares Associated with the Average Ranks of the Errors at the 12 Item Positions, Partitioned into the First Three Trend Components, According to Experimental Condition (Anticipation Method, Eight Ss per Condition)

	Experimental Condition						
	S1	S2	S3	S4	S6	S12	S0
Total χ^2	30.29	37.00	28.39	9.85	34.78	33.52	39.74
Linear Component	.02	.37	2.24	.71	1.13	2.44	13.44*
Quadratic Component	16.54*	21.21*	23.55*	5.29	30.10*	22.19*	21.07*
Cubic Component	2.80	.81	.02	.08	.00	3.09	.06
(Remainder)	(10.93)	(14.61)	(2.58)	(3.77)	(3.55)	(5.80)	(5.17)

* Significant with $\alpha = .005$

data presented in Table 2 represent the total value of χ^2 (with 11 df) associated with the omnibus Friedman test, and the partitioned χ^2 values (each with 1 df) associated with the first three trend components: linear, quadratic, and cubic.

Since 21 tests of hypothesis have been performed (three apiece in seven different experimental conditions), each test was performed with $\alpha = .005$ in order that the error probability for the set of hypotheses would be less than or equal to approximately .10. With $\alpha = .005$, the critical value of χ^2 , with 1 df, is 7.88.

There are at least two striking features of the data in Table 2. In the first place, for all

experimental conditions save one, there is a significant quadratic trend. In conjunction with the means presented in Table 1, it is evident that items toward the middle of the list were generally harder to learn than those at either the beginning or the end.

That this was not true for the S4 condition indicates, perhaps, that the provision of four sentences did something to make the resulting error pattern more uniform, i.e., errors were more evenly distributed among item positions. This finding is further corroborated by the small total χ^2 value of 9.85 which, with 11 df, does not approach statistical significance.

Table 3
Average Ranks of the Transitional Error Probabilities at the 11 Item Transitions, According to Experimental Condition (Recall Method, Eight Ss per Condition)

Condition	Item Transition Number										
	1	2	3	4	5	6	7	8	9	10	11
S1	2.69	3.12	4.56	8.75	8.31	7.38	6.19	5.75	7.88	6.19	5.19
S2	4.50	5.31	6.56	6.94	7.19	5.25	5.50	5.88	7.50	6.06	5.31
S3	3.06	5.00	5.25	5.88	7.12	6.75	5.94	7.19	7.25	5.88	6.69
S4	3.19	4.75	4.69	6.19	7.44	5.62	7.12	7.38	7.25	6.56	5.81
S6	3.44	3.31	5.00	6.75	6.75	7.12	6.44	6.81	5.88	7.69	6.81
S12	3.94	4.12	4.50	6.31	6.69	6.94	6.12	5.69	7.25	7.50	6.94
S0	1.62	4.19	5.81	5.56	7.56	4.56	6.88	7.25	9.38	6.56	6.62

Table 4
Chi-Squares Associated with the Average Ranks of the Transitional Error Probabilities at the 11 Item Transitions, Partitioned into the First Three Trend Components, According to Experimental Condition (Recall Method, Eight Ss per Condition)

	Experimental Condition						
	S1	S2	S3	S4	S6	S12	S0
Total χ^2	29.39	6.46	11.27	13.46	15.68	11.84	30.24
Linear Component	4.66	.24	5.58	6.00	8.96*	8.07*	15.15*
Quadratic Component	14.26*	1.51	3.52	5.12	3.30	1.07	5.34
Cubic Component	1.15	.38	.36	.02	.58	.17	.13
(Remainder)	(9.32)	(4.33)	(1.81)	(2.32)	(2.84)	(2.53)	(9.62)

* Significant with $\alpha = .005$

The second interesting statistic in Table 1 reveals a significant linear, in addition to quadratic, component for Ss in the S0 (o control) condition. The positive value of the weighted sum of coefficients describing the linear trend indicates that errors tended to increase, going across item positions. Such a primacy effect, or the relatively superior recall of initial over final items, when coupled with the significant quadratic component depicts the familiar distribution of serial-position errors typically observed in rote-learning situations.

Recall

For reasons not discussed here, an index in the spirit of Johnson's (1965) transitional error probability (TEP) analysis was considered more appropriate for representing the error profiles obtained under the recall method of testing. Thus, pseudo-TEPs were computed for each S as follows: The number of times that S recalled a particular item (i) in the list up to the trial before criterion was noted (n_{c_i}). It was not required that the item be remembered in its appropriate list position in order for it to be scored as a correct response. After counting S's correct responses to the i th item, the number of *incorrect* responses given by S to the i th + 1 item, or $n(e_{i+1} | c_i)$, was determined. Thus, S's TEP for Transition 1, i.e., between the first and second item, was simply $[n(e_2 | c_1)]/n_{c_1}$ or the number of times that the second item was wrong, given that the first item was correct, divided by the number of times that the first item was correct.

In this manner, 11 different TEPs were produced. As was done in the analysis of serial-position errors, each S's TEPs were rank-ordered across the 11 item transitions. These ranked data were then statistically analyzed according to the previously discussed Friedman (1937) and Marascuilo & McSweeney (1967) techniques.

The distribution of average ranks for TEPs at the 11 transitions, according to experimental condition, may be found in Table 3. A preliminary inspection of these ranks suggests that there are differences in the conditional probabilities of making errors at various item transitions. The total χ^2 statistic and its partitioned trend components are presented for each experimental condition in Table 4. Once again, each test was performed at $\alpha = .005$ with the corresponding critical value of χ^2 , with 1 df, equal to 7.88.

From Table 4 it may be observed that the two largest values of the total χ^2 come from the extreme groups, S1 ($\chi^2 = 29.39$) and S0 ($\chi^2 = 30.24$). Of more interest, however, is that a quadratic component characterizes S1, while a linear component best describes S0. Further, this significant linear trend is found in both S6 and S12 (where many

sentences were employed to provide a verbal context for item pairs or individual items), but not in S1, S2, S3, or S4 (where one or a few sentences were employed to relate three or more of the list's items).

Looking back at the mean ranks in Table 3, one should note that in the S6, S12, and S0 conditions there is a tendency for ranked TEPs to increase going across item transitions. With S1, S2, S3, and S4, however, there is no such increase in mean ranks. For S1, the trend may be observed as first increasing and then decreasing, yielding the significant quadratic trend associated with it. For S2, S3, and S4, one is led to the conclusion that there are only chance deviations from the expected average rank of 6.00.

It is possible that in the less organized conditions (S6, S12, and S0), serial learning under the recall method proceeded in kind of a rote, piecemeal fashion. That is, on the first few trials only the initial items of the list were acquired in their correct serial order, while scrambled bits and pieces from the rest of the list were tacked on as well. With increasing trials, these latter items became more correctly associated in serial order, and then added to the already mastered preceding chunks. Such performance would result in a small proportion of TEPs for early items in the list, and a larger proportion of these errors for items occurring later in the list, which is in accordance with the trend data presented.

Under the assumption that the well organized conditions (S1, S2, S3, and S4) provided various item links throughout the list, it might be predicted that for these conditions TEPs would be approximately equal at each item transition. Specifically, since at each transition in the list, there were connections provided by the preceding item or items, no dramatic increase in the number of conditional errors would be expected at the lattermost transitions. The absence of a linear trend in any of these better organized conditions provided empirical support.

REFERENCES

- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 1937, 32, 675-701.
- JENSEN, A. R. Is the serial-position curve invariant? *British Journal of Psychology*, 1962, 53, 159-166.
- JOHNSON, N. F. The psychological reality of phrase-structure rules. *Journal of Verbal Learning & Verbal Behavior*, 1965, 4, 469-475.
- MARASCUILLO, L. A., & MCSWEENEY, M. Nonparametric post hoc comparisons for trend. *Psychological Bulletin*, 1967, 67, 401-412.
- MEDNICK, S. A. *Learning*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
- MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *American Journal of Psychology*, 1950, 63, 176-185.
- SIMPSON, W. E. Effects of approximation to sentence word-order and grammatical class upon the serial learning of word lists. *Journal of Verbal Learning & Verbal Behavior*, 1965, 4, 510-514.
- SIMPSON, W. E. Errors versus correct responses in the serial learning of word lists. *Psychonomic Science*, 1967, 7, 213-214.
- support of the research by the Graduate Division of the University of California, and the inspiring guidance received from Dr. William D. Rohwer, Jr. The author is indebted to Dr. Donald Hardy and his staff at the Inland Valley Elementary School in Orinda, California, without whose support this study would not have been possible.
2. A more detailed description of the experimental materials and procedures may be obtained from the author upon request.
3. The method selected for analyzing serial-position errors is a potentially useful alternative to that recommended by Jensen (1962), in comparing serial-position curves from different conditions which are not equal with respect to task difficulty.

NOTES

1. This article is based on the author's thesis, submitted in partial fulfillment of the PhD degree at the University of California, Berkeley. The author wishes to acknowledge the financial

Natural language mediation and interitem interference in paired-associate learning¹

JACK A. ADAMS, JOHN S. McINTYRE,² and HOWARD I. THORSHEIM,³ *University of Illinois, Urbana, Ill. 61801*

The effects of interitem interference and natural language mediators (NLMs) on verbal performance were assessed by giving 200 Ss eight trials on a list of 10 paired associates. The NLMs were recorded for each item, each trial. Mediated items had higher performance than rote items. The final items on a trial were subject to more interitem interference and were recalled less well than initial items, with no appreciable difference between rote and mediated items.

Recent studies (e.g., Tulving & Arbuckle, 1963) demonstrated that interference among items of a verbal list is a potent influence on item recall. Adams & Montague (1967) have shown that natural language mediators (NLMs), which are the idiosyncratic associative devices that Ss report they often use to learn paired associates, can reduce the amount of list interference in a RI design. In the present study the Adams-Montague findings are extended to the individual item, with the expectation that mediated items will show less interitem interference.

PROCEDURE

All Ss received eight trials on a list of 10 paired associates. The anticipation method was used with the 4:4 sec rate. There was no break between trials; the 80 items were

presented continuously except for one brief apparatus delay discussed below.

Our procedure for an item was to have S report the response term if he could when the stimulus term alone was showing, as is standard for the anticipation method. The next presentation was the stimulus and response terms of the pair together for knowledge of results and learning, and it was here that we asked S to verbalize his learning mode and speak his NLM aloud if he was using one or to say "Rote" if he was not. The E did not record the NLM verbatim; he checked only whether the item was NLM or rote. Instructions explained mediated and rote learning without emphasizing either one. Three trials were given on a list of five neutral pairs for familiarization.

Ten different starting orders of the list were defined, and there was a different order of the 10 items on each subsequent trial. Furthermore, the logic gave a fixed sequence of 1, 4, 5, 7, 8, 10, 11, 13, 14, and 17 items intervening between the two presentations of items on successive trials. The item which was in Position 9 on Trial $n - 1$ was always in Position 1 on Trial n to give retroactive inhibition (RI) in Amount 1, the item in Position 7 on Trial $n - 1$ was always in Position 2 on Trial n for RI in Amount 4 etc. All items preceding the stimulus slide for a particular item exert a proactive inhibition (PI) effect, and over trials the items contributing to PI steadily increase,