JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 9, 707–715 (1970)

Studies of Learning to Learn
X. Nonspecific Transfer Effects in Free-Recall Learning[1]

LEO POSTMAN, SHEILA BURNS, AND LYNN HASHER
University of California, Berkeley, California 94720

Learning to learn in multiple-trial free recall was investigated. The experiment comprised a training and a test phase. Three methods were used in the training phase: random order of presentation and free recall (RF); constant order of presentation and free recall (CF); constant order of presentation and serial recall (CS). The method in the test phase was always RF. Consistency of output order was manipulated by instructions in both phases. Acquisition of the training lists was faster for CF and CS than for RF but practice gains were comparable for the three methods. Method of prior practice per se did not influence test-list performance; however, there was an interaction of conditions of training and instructional treatments.

This experiment investigated learning to learn (LTL) in multiple-trial free recall (MTFR). The fact that performance in MFTR improves as a function of practice is by now well established (Dallett, 1963; Mayhew, 1967; Rosner, 1970; Tulving, McNulty, & Ozier, 1965). The primary purpose of the present study was to determine how specific the skills are that *S* acquires when he learns successive lists by the free-recall method. Under standard conditions of MTFR the items are presented in a different random order on each study trial and *S* is instructed to recall them in any order he wishes. To what extent will performance of such a conventional MTFR task benefit from prior experience with other methods of practice that place constraints on the order of input, the order of output, or both? The question is of theoretical interest because it bears directly on the characteristics of the habits which are responsible for the observed improvements in free-recall performance. Some of the relevant habits acquired by an experienced *S* may be assumed to be essentially independent of the sequential features of the practice tasks. A case in point is the ability to distribute rehearsal time efficiently so as to maximize the amount of attention given to relatively difficult items. Other habits, particularly those related to the development of subjective units, are potentially tied to a given method of practice. The basis on which items are grouped is likely to be different when the order of input is constant rather than random, and also when recall is serial rather than free. To the extent that grouping operations are task-specific, some decrement in performance would be expected when an experienced *S* encounters a change in the sequential characteristics of the learning task. The absence of such a decrement would indicate that grouping operations reflect a higher-order skill which is readily adapted to a new method of practice.

As in previous studies (cf. Postman, 1970), the logic of standard transfer designs was extended to the analysis of the components of LTL. Accordingly, there was a common test task which was used to evaluate the effects of different training procedures. The conditions manipulated during the training phase were the order of presentation of the words on the study trials and the prescribed order of output on the test trials. The terminal task was always MTFR, with no sequential constraints on either input or output order. Differences in performance on the common test task thus

[1] This research was supported by grant MH 12006 from the National Institute of Mental Health.

could be attributed to the conditions of prior training.

This basic design was expanded for the purpose of examining the conditions of, and the effects of, the increases in the degree of subjective organization (SO) which characteristically occur as a function of practice. Interest centered on three related questions: (*a*) Does the adoption of output strategies favorable to the development of SO enhance the amount of LTL? (*b*) Does *S*'s ability to implement such strategies improve with practice? (*c*) To what extent do output strategies generalize from one type of learning task to another? The exploration of these questions took its point of departure from the assumption that output strategies can be effectively modified by instructions to *S*. This assumption is supported by the finding that the level of SO can be raised substantially by directions to *S* encouraging him to group individual items into higher-order units. Such instructions also serve to accelerate the rises in SO as a function of practice (Mayhew, 1967). Instructions were, therefore, used to manipulate output strategies. The critical instructions were introduced either at the beginning of the training phase or in the test phase. Comparisons between differently instructed groups at successive stages of training were expected to provide information about the relation between changes in SO and LTL. As in the evaluation of the different methods of practice, the manipulations in the test phase were directed toward the assessment of the dispositions developed during the training phase.

Method

Design

The procedure for all experimental groups comprised a training phase and a test phase. Two successive lists were learned by a given method in the training phase. A single list was learned in the test phase. The training lists were learned under one of three different methods of practice: random order of presentation on study trials and free recall on test trials (RF); constant order of presentation and free recall (CF); constant order of presentation and serial recall (CS). When recall was free, two types of instructions were used—a standard (St) version and one encouraging subjective organization (SO) of the items. With the addition of the instructional treatments there were altogether five conditions of training: RF(St), RF(SO), CF(St), CF(SO), CS. The method of practice in the test phase was either RF(St) or RF(SO). Thus, the experimental design was a 5×2 factorial, with five conditions of training and two conditions of testing.

Lists

The learning materials were nine 20-word lists. Each list contained 10 one-syllable nouns and 10 two-syllable nouns. Half the words had ratings of A and the other half ratings of AA in the Thorndike-Lorge frequency count. There were five duplications of first letters in a list. The materials were arranged into three sets of three lists each. Meaningful similarities and associative relations within lists and between the lists in a given set were minimized by inspection. The three sets were used equally often under all experimental conditions. The order of the individual lists within sets was fully balanced, that is, each list occupied the three ordinal positions with equal frequency. In Conditions CF and CS there were two different orders of presentation per list; in Condition RF two different starting orders were used equally often.

Procedure

Practice on each list consisted of six study-test cycles. On study trials the words were presented in the window of a memory drum at a 1.5-sec. rate. The tests were oral and 1 min. long. There was a 3-min. interval between the first and the second list and a 4-min. interval between the second and the third list.

In Condition RF the words were presented in a different random order on each study trial. The orders were subject to the restriction that no item appear in the same serial position more than once and that there be no duplication of sequences of two or more items. In Conditions CF and CS the words were presented in the same order on all study trials.

The instructions always indicated whether the order of presentation on successive study trials would remain the same or would change. The RF(St) and CF(St) instructions informed *S*s that they were free to recall the items in any order they wished. The RF(SO) and CF(SO) instructions urged *S*s to try to group the words in some convenient or useful way and to maintain the groupings from one test to the next. The *S*s were told, however, to feel free to add to the groupings or to change them as they went along; they were asked to recall all the words they could remember even if they had not yet grouped them. The CF(SO) instructions also made it clear that the groupings chosen by *S*s might or might not be independent of the order of presentation. In Condition CS *S*s were requested to reproduce the words in the order in which they appeared.

Full instructions were given at the beginning of the training phase; an abbreviated version of the same instructions was presented before second-list learning and, where appropriate, before third-list learning. Changes in the conditions of presentation and in output instructions were made as explicit as possible prior to the presentation of the test list. The cases in which there was a shift from SO to St instructions warrant a special comment. In these cases the directions in the test phase omitted the request that *S*s attempt to group the items; a statement was added that it was up to *S*s whether they would continue to group the items as before. The conditions in which there was such a shift, for example, from RF(SO) to RF(St), permit an assessment of the effects of a change in instructions *per se*.

Subjects

There were 18 *S*s in each of the ten groups. The *S*s were undergraduate students at the University of California who were naive to verbal learning experiments. Assignment to conditions was in blocks of 10, with one *S* from each condition per block.

Results

Acquisition of Training Lists

The mean total numbers of correct responses given on the six test trials of first-list and of second-list learning are presented in Table 1. The protocols of the CS groups were scored without regard to order. The groups treated alike in the training phase but assigned to different instructional treatments in the test phase have been combined. At this point attention is directed only toward the five conditions of training.

TABLE 1

Mean Total Numbers Correct during Acquisition of First and Second List in the Training Phase

Condition	List 1	List 2
RF(St)	81.0	86.4
RF(SO)	82.0	90.5
CF(St)	85.9	93.5
CF(SO)	88.9	96.3
CS	88.0	95.2

The three methods of practice will be compared first. Performance was significantly better when the order of input was constant (CF and CS) than when it was random (RF), $F(1, 175) = 13.95$, $p < .01$. Conditions CF and CS did not differ reliably, $F < 1$. As compared to conventional MTFR, therefore, a constant order of input facilitated learning regardless of output requirements. Under all methods of practice the second list was learned faster than the first, $F(1, 175) = 102.11, p < .01$. The extent of improvement did not interact reliably with the method of practice, $F < 1$. Thus, the amounts of LTL were comparable for the three tasks. Under conditions of free recall (RF and CF), the instructional treatment failed to produce statistically significant effects. Neither the main effect of instructions nor any of the interactions approached significance. It should be noted, however, that under the RF method the gain between the first and the second list appeared to be greater under SO than under St instructions.

The analyses which follow are concerned with the characteristics of performance that contribute to the gains as a function of practice. The trends in intertrial retention will be considered first. As *S* became more experienced, did recall become more stable, that is, was there an increase in the percentage of items recalled on trial *n* that were given again on trial $n + 1$ (percentage of CC)? These percentages were determined for three pairs of adjacent trials: 1 and 2, 3 and 4, 5 and 6. The measures for the two training lists are summarized in Figure 1. The values have been averaged over the two instructional treatments for RF and CF. The percentages increased, of course, over trials. They were substantially larger during the acquisition of the second than of the first list. In fact, there was only a slight drop between the end of List 1 and the beginning of List 2. The values were clearly higher when the order of input was constant rather than random. Thus, the trends were essentially the same as for the measures of learning. The statistical analysis of these results was confined to the average values of the three CC scores for each list. The conclusions were the same as for the acquisition scores with respect to both the differences among the methods of practice and the gains between the first and the second list.

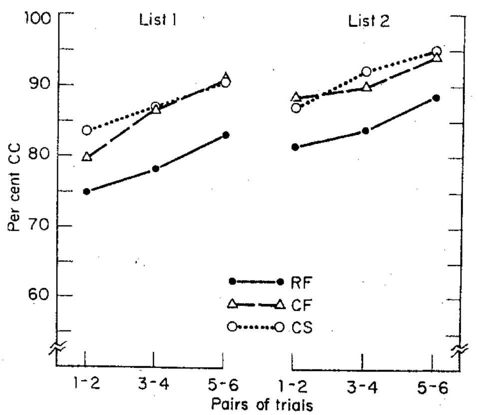

FIG. 1. Mean percentages of CC items (correct on successive trials) during the acquisition of the training lists.

Parallel analyses were carried out for the percentage of items that were missed on trial n but were recalled on trial $n+1$ (NC). The determination of these measures was subject to the difficulty that the base (N) decreased progressively as a function of trials. For some *S*s, N dropped to zero and the percentage became indeterminate. In view of this problem the percentages were calculated for "composite" *S*s. The scores for each composite *S* were based on the pooled observations for three individual *S*s. For this purpose the recall protocols were grouped into six successive sets of three each in accordance with the order in which *S*s had appeared in the laboratory. (In some cases new randomly selected triplets had to be formed in order to eliminate composite *S*s for whom N was zero on one or more trials.) While this procedure handled the strictly computational problem, it did not remove the bias which resulted from the diminishing contribution of fast *S*s to the average scores on the later trials. The mean percentages presented in Figure 2 must, therefore, be interpreted with considerable caution.

The NC values were in general higher for List 2 than for List 1. The differences among the methods of practice were less clear-cut than for the CC percentages. The measures were again higher for CF than for RF, but not appreciably so; CS no longer had an advantage over RF. In the case of the latter comparison the reversal on the terminal pair of trials of first-list learning was unexpected. The pronounced negative acceleration of the curves for second-list learning should also be noted. Differences in opportunities for improvement may be responsible, in part at least, for the apparent deviation of the NC trends from those for acquisition and intertrial retention. Learning was faster under the CF and CS than under the RF conditions. Hence, relatively slow *S*s and difficult items carried greater weight in determining NC

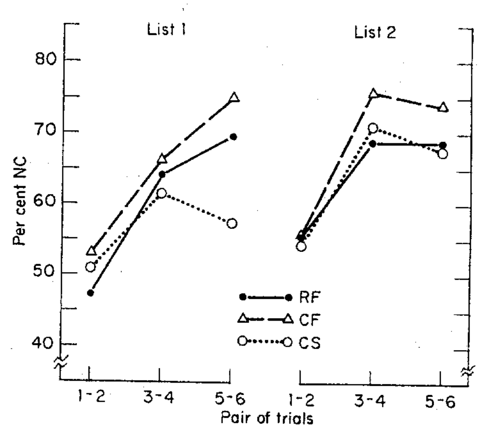

FIG. 2. Mean percentages of NC items (incorrect on trial n and correct on trial $n+1$) during the acquisition of the training lists.

values under the former treatments than under the latter. This bias may be expected to become more important as speed of learning increases. The cases in which there was a lack of improvement in the terminal phase of acquisition may be understood on this basis.

In spite of these uncertainties of measurement the consistently low scores of the CS groups and the relatively small separation between CF and RF suggest that under conditions of constant input *S*s experienced difficulties in adding new items from unfavorable positions. Subjects instructed to reproduce the items in serial order (CS) would be disposed to withhold items they could not place correctly; such a tendency would serve to depress the NC values more than under conditions of free recall (CF). It should be noted that typical serial position curves were obtained for both CF and CS. These curves were comparable, except for the presence of a greater primacy effect under CS than under CF. Thus, it is not possible to specify particular serial positions responsible for the difference in trends between the two procedures. Taken together, Figures 1 and 2 indicate that the advantage of a constant over a random order of input stems primarily from an increase in intertrial retention; differences in the rate at which new items are added on successive trials may be relatively minor or absent altogether. This conclusion must remain tentative in view of the obvious reservations about the validity of the NC scores. (The inherent limitations of this measure are worthy of emphasis.) It is clear, however, that both measures increased as a function of prior practice and thus may be considered as indices of LTL.

Subjective Organization in the Training Phase

The next analysis is concerned with the development of SO in the training phase. The degree of SO was measured in terms of the number of intertrial repetitions (ITRs) of pairs of items in either the same or the reverse order. The difference between the observed number of ITRs and the number to be expected by chance (O – E) was determined for each *S*. These measures were again calculated for trials 1 and 2, 3 and 4, 5 and 6. The mean values for the groups trained under conditions of free recall (RF and CF) are presented in Table 2. The scores increased over trials on a given list. Like the learning measures, the ITRs were higher when the input remained constant (CF) than when it changed from trial to trial (RF). The level of SO was consistently higher for List 2 than for List 1; the amount of increase was comparable for RF and CF. Grouping instructions served to raise the level of SO. The difference between the two instructional treatments developed progressively, however, and was clearly greater during second-list than during first-list learning. This trend was especially pronounced for the RF groups.

The averages of *S*s' three scores on each list were subjected to an analysis of variance. The main effects of the method of practice, type of instructions, and stage of training were all significant at the .01 level, $Fs\ (1, 140) = 23.44$, 13.79, and 94.27, respectively. There was also a reliable interaction of stage of training with instructions, $F(1, 140) = 14.24$, $p < .01$. This interaction reflects the greater increase in ITR scores for the SO than for the St groups.

The conditions tested by serial recall (CS) were not included in this analysis. The instructions to reproduce the items serially were

TABLE 2

MEAN NUMBERS OF ITRs (O-E) UNDER CONDITIONS OF FREE RECALL IN TRAINING PHASE

	List 1 trials				List 2 trials			
Condition	1–2	3–4	5–6	Mean	1–2	3–4	5–6	Mean
RF(St)	.752	1.710	2.693	1.718	.880	2.725	3.250	2.285
RF(SO)	.743	2.092	2.708	1.848	1.810	4.044	6.044	3.966
CF(St)	.966	2.490	4.221	2.559	1.430	4.067	6.338	3.945
CF(SO)	1.370	3.314	5.713	3.466	2.380	6.373	8.584	5.779

carried out by all *S*s. Thus, the correlation between the order of input and the order of output rapidly approached ceiling during first-list as well as second-list learning.

Acquisition of Test List

Figure 3 shows the curves of test-list acquisition. In order to bring out as clearly as possible the effects of the method of prior practice, the results have been summed across the

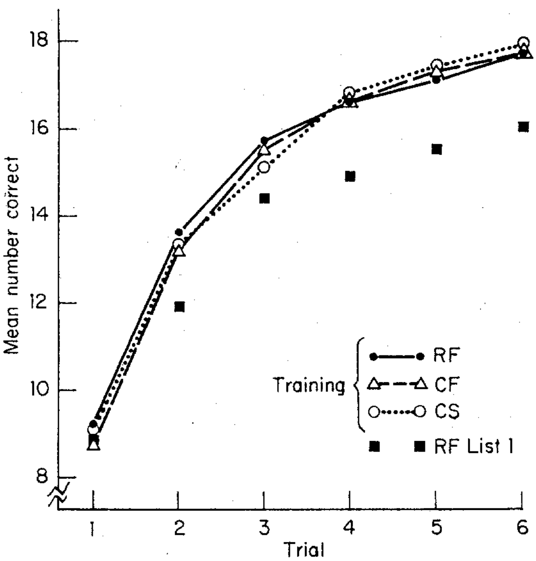

FIG. 3. Curves of test-list acquisition under different conditions of prior practice. Unconnected solid squares are first-list scores of the RF-RF groups.

different instructional treatments. It is obvious that the method of practice *per se* had no influence on performance in the test phase: a change in method did not retard performance in the final list. When the method remained the same (RF-RF), the gains achieved in the training phase were maintained. In order to show the amount of improvement during the experimental session, the first-list scores of the groups which continued learning by the same method have been included in Figure 2 (unconnected squares). Clearly the slope of the learning curve was substantially steeper during the acquisition of the third than of the first list. When there was a shift from a constant to a random order of presentation (CF-RF and CS-RF) the level of performance declined but did not fall below that of the groups for which there was no change in procedure.

TABLE 3

MEAN TOTAL NUMBERS CORRECT DURING ACQUISITION OF THE TEST LIST

	Test		Overall
Training	RF(St)	RF(SO)	mean
RF(St)	92.9	85.8	89.4
RF(SO)	89.7	90.8	90.2
CF(St)	83.2	87.9	85.6
CF(SO)	96.6	88.4	92.5
CS	89.3	89.9	89.6
Overall mean	90.3	88.6	

To permit a more detailed assessment of the different conditions of transfer Table 3 presents the mean total numbers correct in test-list learning for each of the treatment combinations. Interest now centers on the effects of the instructional manipulations. There was no reliable overall difference between the groups given the two types of instructions in the test phase, $F < 1$. In the case of the groups which had practiced the prior lists under conditions of free recall, the only significant source of variance was the higher-order interaction, Method of practice × Training instructions × Test instructions, $F(1, 170) = 6.86$, $p < .01$.[2] The nature of the interaction may be summarized as follows: When the input conditions remained the same (RF-RF), a change in instructions decreased the efficiency of test-list learning; when the input conditions changed (CF-RF), a shift in instructions was beneficial. Thus, RF(St)-RF(St) was superior to RF(St)-RF(SO), whereas the difference

[2] A check was previously made on the comparability of the groups treated alike in the training stage but assigned to different instructional treatments in the test stage. In an analysis of variance of the second-list scores none of the sources of variance associated with the treatment in the test stage approached significance, all $Fs \leqslant 1.29$.

between CF(St)-RF(St) and CF(St)-RF(SO) was in the opposite direction. Output strategies are necessarily geared to the conditions of input. Changing one of these conditions of performance but not the other reduced the usefulness of the operations developed earlier and at the same time made it more difficult to develop new modes of attack on the test task.

The CC and NC percentages for test-list learning are shown in Figure 4. The measures have again been averaged across instructional treatments and thus are directly comparable

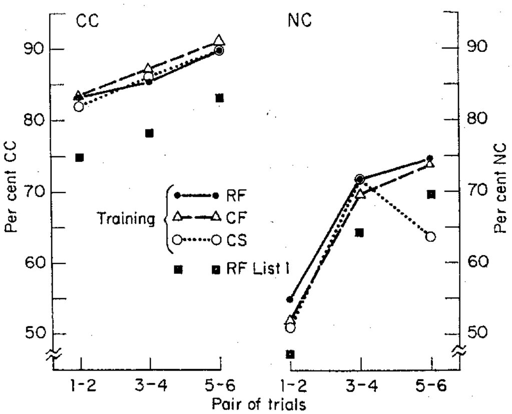

FIG. 4. Mean percentages of CC and NC items during the acquisition of the test list. Unconnected solid squares are first-list scores of the RF-RF groups.

with the acquisition scores presented in Figure 3. The CC values were uniformly high and reflected nearly perfect intertrial retention. Like the measures of test-list acquisition, the percentages were essentially the same under the three methods of prior practice. The substantial gains of the groups trained and tested by the same method are again clearly in evidence. The improvement was somewhat less pronounced for the NC measures, but in view of the limitations of this index discussed earlier this conclusion must be viewed with caution. An intriguing feature of the NC curves is the sharp terminal decline for the groups trained by the CS method. This reversal mirrors the drop observed in first-list learning (see Figure 2). The possibility suggests itself that there was a residual effect of the tendency to "build up" the list by reproducing items serially which made the addition of the last few items difficult. There is, however, no apparent independent support for this speculation.

Subjective Organization in the Test Phase

The ITR scores for the test phase are summarized in Table 4. The terminal instructions proved effective: explicit directions to group the items again increased the level of consistency in the order of output. The overall difference between the two conditions of testing was significant, $F(1, 170) = 8.43$, $p < .01$. For the groups tested by free recall throughout the experimental session (all conditions except CS), the main effect of the instructional treatment during the training phase was likewise reliable, $F(1, 170) = 7.54$, $p < .01$. On the average, then, the ITR scores in the test phase

TABLE 4

MEAN NUMBERS OF ITRS (O-E) IN TEST-LIST LEARNING

	Test Trials							
	RF(St)				RF(SO)			
Training	1–2	3–4	5–6	Mean	1–2	3–4	5–6	Mean
RF(St)	1.661	3.897	5.528	3.695	1.394	3.855	5.266	3.505
RF(SO)	1.546	3.440	4.904	3.297	1.964	5.140	6.379	4.494
CF(St)	.749	2.005	2.371	1.708	1.793	4.010	5.367	3.723
CF(SO)	2.091	5.200	6.431	4.574	1.443	3.827	7.202	4.157
CS	1.291	2.172	3.323	2.262	1.627	4.393	6.780	4.267

were higher when training was under SO instructions than when it was under St instructions. However, the higher-order interaction, Method of practice × Training instructions × Test instructions, was again significant, $F(1, 170) = 7.27$, $p < .01$. The basic finding was the same as for the measures of learning. A change in instructions was disrupting when the input conditions remained the same but tended to have a favorable influence when the input conditions changed as well. After RF training, when there was no change in input conditions, a shift from St to SO instructions produced a slight detrimental effect, and that from SO to St a very large one. After CF training, on the other hand, the corresponding comparisons show a large and a small positive effect, respectively. It can be seen that the interaction was superimposed on the direct influence of the test instructions. The effectiveness of the latter was clearly in evidence after CS training, that is, when there was a change in both input and output conditions. Reference to Tables 2 and 4 shows that there were substantial progressive increases in SO when both the method of practice and the instructions remained constant, that is, for RF(St)-RF(St) and RF(SO)-RF(SO).

Discussion

The results of the experiment add to the evidence for nonspecific transfer in MTFR. Some of the characteristics of performance that reflect the impact of prior practice have been identified: there were increases in intertrial retention, in the ability to add new items after an input trial, and in the measures of SO. It is likely that under all conditions *S*s learned to make optimal use of the available study time and to regulate their output on test trials. The differences observed in the training phase point to the importance of developing interitem associations and grouping individual words into higher-order units. Performance was clearly better when the order of presentation on study trials was constant rather than random, that is, when consistent sequences were provided in input. (This result agrees with the findings of several recent investigations: Jung & Skeebo, 1967; Lachman & Laughery, 1968; Mandler & Dean, 1969.) Explicit directions to group the items significantly increased the degree of output consistency and also tended to raise the level of performance, although not reliably so.

The measures of acquisition in the test phase show that the habits and skills responsible for the nonspecific transfer effects were readily, and indeed fully, generalizable from one method of practice to another. In particular, a shift from a constant to a random order of input, or from serial to free recall, did not in and of itself reduce the practice gains. To the extent that LTL was independent of the task used in training, *S*s acquired higher-order dispositions which were readily adaptable to new experimental requirements. Thus, there were no overall differences in the terminal phase as a function of the method of prior practice because experienced *S*s, regardless of the particular training procedure, were disposed to approach the common test task in similar ways.

Transfer from one method of practice to another requires flexibility in the application of the skills acquired in the course of training. Thus, when there was a shift from CS to RF, the experienced *S* had to be ready to give up "seriation" (Mandler & Dean, 1969) and to group nonadjacent items on the basis of associative and categorical relations. The pattern of scores in the terminal phase indicates that the generalization of learning skills from one task to another did not occur automatically but reflected sensitively variations in the experimental context. This conclusion is based on the higher-order interaction of instructional treatments and methods of practice in the training and in the test phase. The pattern of change in these factors influenced the measures of subjective organization as well as the speed of test-list learning. This interaction suggests that a partial change in the definition of the experimental requirements—in either the conditions of input or the instructions, but not in both—may have prevented *S* from fully applying the skills which he had developed during the training phase. Thus, it may not be paradoxical after all that terminal performance was optimal both when the training and the test conditions were strictly continuous and when they were clearly differentiated. In the former case *S* could continue doing what he had learned to do efficiently; in the latter case he could focus sharply on the new as distinct from the old requirements and act accordingly. The sensitivity of *S* to *E*'s definition of the task is brought out clearly by the fact that the omission, as well as the addition, of instructional statements in the terminal phase appeared to have an appreciable effect on performance.

The question of whether output strategies can be manipulated by instructions was again answered in the affirmative by the results obtained in the training phase. Increases in SO induced by instructions had, however, only a marginal and unreliable influence on speed of learning. Some dissociation of the measures of learning and of output consistency was also found in the test-list scores. Specifically, the main effects of instructional treatments (both training instructions and test instructions) were significant for the ITR scores but not for the measures of acquisition.

These findings are consistent with the view that regularities in the order of output provide incomplete and fallible indices of the degree of organization. While imperfectly related to functional organization, order of output is a characteristic of performance which is under *S*'s control and as such is subject to instructional manipulation. It appears that *S* can change the pattern of his test responses more readily than he can improve the efficiency of his organization of the list.

The pattern of the SO scores in the test phase exhibited the same higher-order interaction as was found for the measures of acquisition. The *S*s' readiness to utilize their newly acquired organizational skills depended on the nature of the transition from the training phase to the test phase. Taken together, the measures of performance in the terminal phase illustrate the joint determination of non-specific transfer by the conditions of prior training on the one hand and the definition and requirements of the test task on the other.

References

Dallett, K. M. Practice effects in free and ordered recall. *Journal of Experimental Psychology*, 1963, **66**, 65–71.

Jung, J., & Skeebo, S. Multitrial free recall as a function of constant versus varied input orders and list length. *Canadian Journal of Psychology*, 1967, **21**, 329–336.

Lachman, R., & Laughery, K. R. Is a test trial a training trial in free recall learning? *Journal of Experimental Psychology*, 1968, **76**, 40–50.

Mandler, G., & Dean, P. J. Seriation: Development of serial order in free recall. *Journal of Experimental Psychology*, 1969, **81**, 207–215.

Mayhew, A. J. Interlist changes in subjective organization during free-recall learning. *Journal of Experimental Psychology*, 1967, **74**, 425–430.

Postman, L. Experimental analysis of learning to learn. In G. H. Bower and J. T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 3. New York: Academic Press, 1970.

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

Tulving, E., McNulty, J. A., & Ozier, M. Vividness of words and learning to learn in free-recall learning. *Canadian Journal of Psychology*, 1965, **19**, 242–252.

(Received June 30, 1970)