

Learning a list for free recall: Selective reminding versus the standard procedure

COLIN M. MacLEOD

University of Toronto, Scarborough, Ontario, Canada

Almost all free-recall experiments involve an acquisition phase in which the set of to-be-learned items is presented repeatedly in its entirety. Two experiments are described that pit this *standard* procedure against the *selective reminding* procedure (Buschke, 1973), wherein only items incorrect on trial n are presented again on trial $n+1$. The two methods were virtually identical in terms of (1) trials to reach an acquisition criterion of one errorless recall, (2) number of items correct on each recall trial during acquisition, (3) number of items correct on a 3-week delayed-recall test, (4) number of items correctly relearned after the 3-week retention interval, (5) extent of subjective organization on all retention tests, and (6) pattern of errors over all test trials. These results generalized to both single-category and random lists of 20 words, and were thoroughly consistent across the two experiments. However, the standard procedure required that many more items be presented over trials, and more subjects failed to learn the list within 1 h using the standard procedure. Because the selective reminding procedure is more efficient and has no apparent drawbacks, it is recommended that memory researchers switch to selective reminding as the preferred way to teach a subject a free-recall list.

"In a typical *free-recall* experiment, subjects are given a list of items and are later asked to recall as many as possible. Usually from 20 to 40 items are presented, one at a time. Recall may be oral or in writing. Subjects are instructed to recall as many words as they can, without regard to the order in which the items have been presented. The recall period varies from 30 seconds to 2 minutes. Order of presentation is randomized from trial to trial" (Kintsch, 1970, p. 10). This is the *standard* procedure for teaching a list to a subject in a free-recall experiment, and has been used in hundreds of learning and memory experiments.

Probably, the standard procedure derives from use of the memory drum as the device for item presentation. With the list typed on a continuous roll of paper, all of the items had to be shown to the subject on each study trial. Although a different order could be used from trial to trial (by preparing several rolls in advance), study trials could not be tailored to the individual subject's performance on a preceding test trial. Even when the list is constructed so that each item is on a different card or slide, as is more common today, it is still conventional to display all of the items on every trial, with only the order changing.

Twelve years ago, Buschke (1973) introduced another technique for conducting the acquisition phase of a free-

recall experiment. He called this the *selective reminding* procedure. "In essence, selective reminding simply involves the selective presentation on each recall trial of only those items which were not recalled on the immediately preceding trial" (Buschke, 1973, p. 543). Although Buschke was interested in this procedure mainly for its value in studying subjective organization of the list (e.g., Buschke, 1976; Fuld & Buschke, 1976), selective reminding can be thought of as a general alternative to the standard procedure. Yet almost all investigators continue to use the standard procedure in free-recall learning.

In fact, Buschke's (1973) procedure was not unprecedented. In the paired-associate literature, the "drop-out" technique (see Battig, 1965) parallels selective reminding. Under both procedures, items correct on trial n are removed on trial $n+1$. The argument is that dropping out correct items speeds up study trials because fewer items are presented, while it reduces overlearning of items learned on early trials. Indeed, the all-or-none versus incremental debate of 25 years ago (see Postman, 1962; Rock, 1957) relied partly on data contrasting this procedure with the standard, whole-list procedure. Unlike selective reminding in free recall, however, the drop-out technique has seen continued use in the paired-associate literature (e.g., MacLeod, 1976; Nelson, 1971).

In the computer-assisted instruction literature, techniques such as drop-out and selective reminding have been referred to as "response sensitive" (see Atkinson & Paulson, 1972). An issue of considerable interest in developing instructional programs was how to optimize presentation of to-be-learned materials. Thus, Karush and Dear (1966) maintained that selective reminding was the optimal presentation scheme in the context of an all-or-none learning model. In teaching children to spell, Lorton

This research was supported by Natural Sciences and Engineering Research Council of Canada Grants A7459 and E6532. For their assistance in collecting and scoring the data, I thank Randal Clark, P. J. Lootsteen, and Victoria Prince; for their valuable advice, I thank Robert Bjork, Bert Forrin, Arthur Glenberg, and Paul Muter. Requests for reprints should be addressed to Colin M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada M1C 1A4.

(1972; described in Atkinson & Paulson, 1972) contrasted the standard, whole-list procedure with a version of the selective reminding procedure. Although children made somewhat more errors during acquisition under selective reminding, their retention performance was actually a little better. However, both differences were small—in the 10% range—and some features of Lorton's design make it difficult to generalize to the laboratory list-learning situation.

Selective reminding actually does have some precedent in the free-recall literature. To study optimization of learning, Murdock, Penney, and Aamiry (1970) used a technique wherein all items were presented on each study trial, but subjects were informed of which ones they had already recalled correctly. One group of subjects was shown the words recalled on trial n at the beginning of the whole list on trial $n+1$, whereas the other group saw their recalled words at the end of the whole list on trial $n+1$. Small but reliable differences were observed between the performance of these two groups and that of the control group with the standard randomly reordered presentation. However, in a later study designed to clarify certain aspects of the earlier one, Murdock, Anderson, and Ho (1974) found no evidence that the reordering procedure affected performance.

Like selective reminding, Murdock's reordering procedure has seen little subsequent use. Perhaps the special purposes that Buschke and Murdock offered for introducing these two selective procedures into free recall have led investigators to overlook their possibilities as more general procedures. But are we better off staying with the standard procedure, or do the selective procedures offer some advantages that warrant switching to them? This question is particularly important now in view of the widespread use of computers, which make item selection so easy to implement. Only an empirical comparison of techniques can provide an answer.

Theoretically, there are several ways in which the two procedures might differ. First, because an item correctly recalled on a particular test trial is not repeated on the subsequent study trial, selective reminding might result in less overlearning of items acquired early. This would be consistent with the logic of the drop-out technique in paired-associate learning. Viewed from the other perspective, the list might be learned more fully under the standard procedure. A delayed retention test would provide one way to evaluate any difference in degree of original learning.

Second, because all items recur on every study trial in the standard procedure, it seems reasonable that recall of particular items might be more consistent under this procedure. Selective reminding does not refresh correctly recalled items, potentially leading to fluctuations over successive test trials (see Tulving, 1967; Nelson & MacLeod, 1974). One way to examine intertrial consistency would be via the amount of clustering of items, or subjective organization of the list, over trials. Does the standard procedure lead to greater list organization?

A third apparent difference between the two procedures relates to the extent of retrieval information each provides during acquisition. In a way, selective reminding tells the learner which items are posing the greatest difficulty, whereas the standard procedure gives no diagnostic assistance at all. Following the logic used in computer-assisted instruction, such retrieval information might permit a subject's learning by selective reminding to reach acquisition earlier by selectively rehearsing difficult items. Trials to criterion and number of items correct on each acquisition test trial provide possible indices of this retrieval advantage. Subjective organization measures also might assist in evaluation.

The two experiments reported here compare selective reminding to the standard procedure. Because the two acquisition procedures might differ in some ways but not in others, a variety of dependent variables should be examined. In both of these experiments, trials to acquisition, number correct on each successive test trial, trial-to-trial output order (subjective organization), and total number of items presented during study were recorded. These should characterize acquisition quite well. To characterize retention, a 3-week delayed test was administered, and subjects were given a single relearning study and test trial. Both correct and error responses were analyzed on all recall trials.

These experiments are, admittedly, exploratory in nature. The potential differences just outlined will be evaluated, but the overall aim is to compare the methods in as diverse a set of ways as possible. Also, because some of the comparisons that fail to show differences will be noteworthy, acceptance of the null hypothesis becomes a concern. To help alleviate that concern, Experiment 2 is in large part a replication of Experiment 1, demonstrating a highly consistent pattern. Toward the same end, power analyses are included when feasible. Emphasis is placed primarily on comparing *methods*, although theoretical issues relating to the comparison also will be raised. There are two principal questions. First, do these acquisition procedures differ in any way(s)? Second, what theoretical processes might underlie any differences that are observed?

EXPERIMENT 1

In this first study, the two acquisition procedures were pitted against each other, and a large set of dependent measures was collected. To increase the generalizability of the results, two different types of list were used. One list consisted of 20 items from a single category, as in Buschke (1973); the other was made up of one item from each of 20 different categories, as in Fuld and Buschke (1976). It seems important to know whether two procedures that are nominally *general* might interact with different types of material. However, the most fundamental question addressed here is whether the two procedures affect performance differently and, if so, in what way(s) they differ.

Method

Subjects. Forty-seven undergraduates at the Scarborough Campus of the University of Toronto volunteered to participate in the study. They were assigned to the four experimental groups at random. Of the 47, 3 failed to return for the delayed retention test, and 8 were unable to reach the acquisition criterion within 1 h. The remaining 36 subjects were divided equally over the four groups.

Materials. Two different lists were used as study materials to permit greater generalizability of the results, and to allow comparison with Buschke's (1973) results with the selective reminding procedure. One list, used by Buschke, was made up of 20 animal names, all from the ANIMAL category in Battig and Montague (1969); the other list, used by Fuld and Buschke (1976), was made up of one word from each of 20 different categories in Battig and Montague. These are called the ANIMALS and RANDOM lists, respectively. The words were typed individually on 3×5 in. index cards in capital letters for use in the study sequences.

Design. A 2×2 between-subjects design was used. The factors were acquisition procedure (standard or selective reminding) and list type (ANIMALS or RANDOM). Nine subjects were assigned at random to each of the four groups.

Procedure. In the standard acquisition procedure, the subject saw all 20 words in the list individually for 3 sec each during study. The experimenter used audible clicks recorded on a tape recorder to pace item presentation. Immediately following a study trial, the subject was given 3 min to write down as many of the words as possible. As soon as the recall interval was over, the entire list was presented again in a different random order, followed by another test interval. Study and test cycles alternated until the subject correctly recalled all 20 words on a single test. No feedback was provided, and the subject was not told the acquisition criterion.

The selective reminding acquisition procedure followed the same basic pattern, but with one major change. Rather than studying all 20 words on each study trial, the subject was shown only those items that had not been recalled correctly on the immediately preceding test trial. Thus, if on test trial n the subject recalled 13 list items, on study trial $n+1$ only the remaining 7 items were presented. As in the standard procedure, study items were always presented in a new random order. The subject was instructed to recall as many of the 20 words as possible on each test trial, not just those words most recently studied. In all other respects, the procedures for the two types of acquisition were identical.

Three weeks after acquisition, the subjects were telephoned and asked to return for "another experiment." To prevent rehearsal during the retention interval, they had not been told of this second session during acquisition. When they returned, they were given a single retention test for the list, followed by a single relearning study and test trial on the entire list. The relearning procedure was identical to the first trial of original learning. Both tests were restricted to 3 min. The subjects were debriefed after the second session.

Results and Discussion

The major question was whether the two different acquisition procedures, standard and selective reminding, would lead to any observable differences in memory performance. To answer this question, a rather wide set of dependent variables was considered because the possibility existed that differences between the procedures might be restricted to only some measures.

Acquisition. Of the 3,851 individual words recalled during acquisition (collapsing over groups), only 61 (1.6%) were incorrect. Half of these were extralist intrusions, and half were repetitions. Most errors were on the first few test trials. Although the infrequency of errors prevents statistical analysis of them, there did not

appear to be any differential error rate over the experimental conditions.

The most obvious measure of acquisition performance is the number of trials to criterion, in this case one errorless recall of the entire list. These data are displayed at the top of Table 1. A 2×2 analysis of variance, using acquisition procedure and list type as the two independent variables, revealed no significant effects. Only the value for acquisition procedure [$F(1,32) = 1.68, p > .20$] exceeded unity. Using the widely accepted standard power of .8 as a criterion (see Keppel, 1982, p. 72), this analysis could have detected a difference between means for the two procedures of 1.98 trials. As Table 1 affirms, the two procedures are quite comparable, regardless of type of list studied.

Figure 1 displays another acquisition variable—average number of items correctly recalled on each acquisition trial.¹ This provides a rough measure of the rate of approach to the criterion of list acquisition. A 2×2 analysis of variance using number correct on each trial (collapsed over subjects) confirms the apparent similarity in rate of acquisition over groups. All of the F values—for procedure, list, and procedure \times list—were less than unity. Again assuming a power of .8, this analysis could have detected a difference between means of 2.38 items. In addition to confirming the above results for trials to reach the acquisition criterion, this analysis also suggests that, if one were to conduct only a certain fixed number of acquisition trials, the methods would still be quite comparable.

Despite the surface comparability of the two methods, there might be important dissimilarities beneath. Put another way, although global measures such as number of trials or number of correct items per trial may not differ, there could still be qualitative differences within the trials. One widely used tool for examining within-trial structure is the extent of clustering (Bousfield, 1953) or

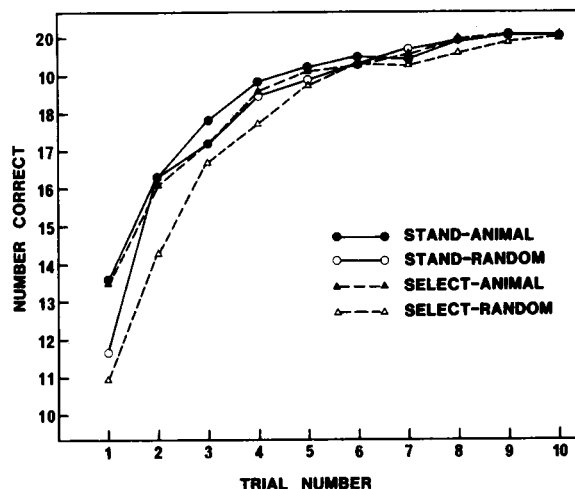


Figure 1. Experiment 1: Number of items correctly recalled on each successive test trial during acquisition as a function of list type and acquisition procedure.

Table 1
Acquisition: Mean Number of Trials to Criterion and
Mean Number of Items Presented as a Function of
Acquisition Procedure and List Type

Type of List	Acquisition Procedure	
	Standard	Selective Reminding
	Trials to Criterion	
ANIMALS	5.33	6.55
RANDOM	6.22	6.77
	Total Number of Items Presented	
ANIMALS	106.66	36.77
RANDOM	124.44	43.77

subjective organization (Tulving, 1962) employed. This refers to the degree of consistency in recall order across test trials. To measure subjective organization, Murphy and Puff (1982, pp. 120-123) recommended Tulving's (1962) index, which can be given as $SO = O(ITR)/(C + 1)$, where $O(ITR)$ represents the number of observed intertrial repetitions of item pairs (without regard to order in the pair) and C represents the number of common items recalled on successive trials. For n trial, there will be $n - 1$ SO scores.

Median SO scores across the first nine acquisition trials (for which all groups provided data) demonstrate a similar pattern: (1) standard-random—.18, .22, .24, .21, .20, .26, .15, and .35; (2) standard-animals—.18, .15, .20, .21, .24, .28, .20, and .30; (3) selective reminding-random—.20, .20, .21, .25, .32, .37, .39, and .38; and (4) selective reminding-animals—.15, .22, .26, .37, .33, .31, .34, and .55. Subjective organization increases across trials in all four groups. Indeed, the specific preferred item pairings are actually very similar across subjects and groups within each list. Although there appears to be slightly more extensive clustering under the selective reminding procedure, this is true only on later acquisition trials. However, fewer observations contribute to the estimates on later trials (because more subjects have reached criterion), so these data are not as stable as those for earlier trials. At this point, then, all that can be said is that there is little evidence in support of any within-trial organizational difference due to acquisition procedure. This was pursued in Experiment 2, in which the larger sample size permitted a stronger evaluation of any differences in subjective organization.

In addition to number of trials to criterion, Table 1 displays the mean number of items presented over all acquisition trials to each subject. It is here the only reliable difference between the procedures emerges. A 2×2 analysis of variance revealed only one significant effect, that of acquisition procedure [$F(1,32) = 56.24$, $p < .001$]. The main effect of list type was nonsignificant [$F(1,32) = 1.52$, $p > .20$], as was the interaction ($F < 1$). Quite clearly, a subject learning a list via the standard acquisition procedure must study many more items over trials than must a subject learning via selective reminding. Consequently, acquisition can take much

longer for subjects who learn under the standard procedure.

One other observation highlights this difference in relative difficulty of acquisition using the two procedures. The eight subjects who were unable to learn their list in 1 h were distributed across groups in the following way: two, selective-random; three, standard-animal; and three, standard-random. That the random list is somewhat harder to learn is not surprising; indeed, it is suggested by the means throughout the acquisition analyses. What is interesting is the fact that six of the eight subjects who failed to learn the list were in the standard condition. Although only weak evidence, this observation is consistent with the claim that the standard acquisition procedure may be somewhat more difficult than the selective reminding procedure. Even if it is not more difficult, the standard procedure will necessarily take longer because more item presentations occur at study. Consequently, time restrictions on acquisition periods will have more influence with the standard procedure than with the selective reminding procedure, as observed here.

To summarize the acquisition data, there appears to be no disadvantage for selective reminding compared with the standard procedure in terms of rate of approaching list acquisition or number of trials to acquisition. Degree of subjective organization in recall also looks similar under the two procedures, although a stronger test is required. However, a subject who learns via the standard procedure must view many more presentations of the items than must a subject who learns via selective reminding, and these additional presentations necessarily consume more study time. Taken together, these results suggest that the selective reminding procedure is a more efficient acquisition procedure without any cost relative to the more widely used standard procedure.

Retention. These acquisition analyses are informative, but they do not tell the whole story. What about retention subsequent to acquisition? In most memory research, this is of at least as great concern as acquisition. How might these two procedures differ in their effects on retention? In the introduction, two possibilities were raised. On the one hand, the additional presentation in the standard procedure might result in better long-term retention due to overlearning. On the other hand, the specific information provided by selective reminding about which items are most difficult may help to structure encoding and retrieval of the list. It was to test these speculations that the 3-week delayed retention data and relearning data were collected.

As in acquisition, the subjects made very few recall errors on the delayed tests. Only 19 of the total of 533 words recalled by all subjects on the delayed retention test were incorrect, constituting an intrusion rate of 3.2% and a repetition rate of 0.4%. There were only four errors on the 688 words recalled during relearning, 2 intrusions and 2 repetitions. Basically, neither procedure seems to promote errors.

The means for the delayed retention test are shown in Table 2. A 2×2 analysis of variance of the delayed-recall data revealed only one significant effect, that of list [$F(1,32) = 7.95, p < .01$]. The retrieval cue value of all of the items' being from a single category clearly benefits recall of the ANIMALS list, independently of acquisition procedure. The F values for the main effect of procedure and for the interaction were both less than unity. Once again using a power of .8, this analysis could have detected a difference of 3.42 items. Although the detectable difference here is not as small as that for acquisition (because variance is greater at delayed retention), this analysis nonetheless displays reasonable sensitivity. Thus, whether a list is acquired by the standard procedure or by selective reminding, it is remembered about equally well after a long retention interval.

Subjective organization is quite consistent in the four groups on the delayed test, as measured by pair repetitions from the final acquisition test trial to the retention test. The median SO scores again were very similar over groups: (1) .19 in the standard-random group, (2) .24 in the standard-animals group, (3) .33 in the selective reminding-random group, and (4) .22 in the selective reminding-animals group. Most of the clustering appears to derive from the same item pairs in each list over all subjects and groups.

Unfortunately, the relearning data suffer from a ceiling effect, making any conclusion from them risky. A 2×2 analysis of variance of the data shown at the bottom of Table 2 revealed a nonsignificant main effect of acquisition procedure ($F < 1$), but a marginally significant effect of list [$F(1,32) = 3.60, .10 > p > .05$]. Again, it appears to be easier to relearn the one-category list, although this is qualified by a significant interaction [$F(1,32) = 4.70, p < .05$]. Probably, though, the absence of variance in the group using the standard procedure to learn the ANIMALS list accounts for most of this result, and caution should be exercised in interpreting it. A longer retention interval would permit examination of the relearning data unclouded by this ceiling effect, but at the risk of a floor effect on the delayed-recall data.

To summarize the retention data, there seems to be no evidence of any detrimental effect of selective reminding compared with the standard acquisition procedure. This complements the findings for acquisition.

Table 2
Retention: Number of Items Correct on the 3-Week
Delayed Retention Test and on the Test Following
Relearning as a Function of Acquisition
Procedure and List Type

Type of List	Acquisition Procedure	
	Standard	Selective Reminding
Delayed Retention Test		
ANIMALS	16.88	15.00
RANDOM	12.77	12.44
Test Following Relearning		
ANIMALS	20.00	18.77
RANDOM	18.33	18.88

EXPERIMENT 2

Perhaps surprisingly, the only difference between the two acquisition procedures in Experiment 1 was in the number of items studied, with selective reminding subjects seeing many fewer. The result was less time to reach acquisition for the selective reminding group, despite equivalence on all other measures of acquisition and retention. The practical implication is that selective reminding is a more efficient acquisition procedure than is the standard technique. To confirm this, it seemed important to replicate the findings of Experiment 1. This was the major purpose of Experiment 2. Here, only the random list was used, with a considerably larger sample size to increase confidence in the results. The only other changes were that Experiment 2 was conducted under computer control and that efforts were made to eliminate the ceiling effect in relearning and to obtain a clearer picture of subjective organization. Discussion of theoretical issues is postponed until Experiment 2 has been presented.

Method

Subjects. Forty-seven undergraduates at the Scarborough Campus of the University of Toronto volunteered to participate for course credit in introductory psychology. They were assigned to conditions at random and took part in the sessions individually. The data of eight subjects are not included: Two subjects in the selective reminding condition did not follow instructions during acquisition, one subject in the standard condition did not complete acquisition in 1 h, one set of data was lost due to a power failure, and four subjects failed to return for the second session.

Of the 39 subjects who completed the experiment, 20 were in the selective reminding condition and 19 were in the standard condition. The second session was to occur exactly 3 weeks later, but 1 day in either direction was allowed when rescheduling problems arose. The mean retention interval was 21.11 days.

Materials. The experiment was carried out under the control of an Apple II+ microcomputer. Because the pattern of results was similar for both lists in Experiment 1, and to provide more data for the analysis of subjective organization, only one acquisition list was used in Experiment 2. This was the Fuld and Buschke (1976) list, made up of 1 word from each of 20 different categories in Battig and Montague (1969). A further 10 words were selected, 1 from each of 10 previously unused categories in Battig and Montague, to be comparable to the acquisition set. These were used in the relearning phase of the second session, as described below.

Procedure. Both the standard procedure and the selective reminding procedure were carried out in very much the same way as in Experiment 1, except that item presentation during study was at the center of a video screen under control of the microcomputer. At test, the screen displayed the number "1" and a question mark, and the subject was to type in the first list word he or she remembered. This was repeated for the numbers 2 through 20. If unable to recall 20 words, the subject was to press the RETURN key enough times to go on to the next study trial. Responses remained visible on the monitor so that the subjects could avoid repeating items on a single recall test and because this was most analogous to the written recall procedure of Experiment 1. In all other respects, acquisition proceeded precisely as in Experiment 1.

Three weeks later, the subjects were telephoned and asked to return for the previously unannounced retention tests. Without further study, a single free-recall test of the 20 items was conducted at the subjects' pace. Then a single relearning study trial was administered. Rather than relearning the entire original 20-item list (as in Experiment 1), the subjects learned a list made up of 10 of the original items and 10 new items. This change was introduced

to make relearning more difficult, with the hope of eliminating the ceiling effect observed in Experiment 1. Half of the subjects from each acquisition condition studied one set of 10 original items plus the 10 new items; the other half of the subjects studied the remaining 10 original items plus the 10 new items. The relearning list was a different random order of the 20 items for each subject. In all other respects, the second session of Experiment 2 was conducted in the same way as the second session of Experiment 1.

Results and Discussion

As was the case in Experiment 1, a quite large set of dependent variables was considered in comparing the two acquisition procedures in Experiment 2. Do the standard procedure and the selective reminding procedure produce any observable differences either at the time of acquisition or on a delayed retention test?

Acquisition.² The first measure of list learning to consider is the number of trials to criterion, here one errorless recall of the entire list. A one-way analysis of variance demonstrated that the means for the selective reminding condition (5.85 trials) and for the standard condition (5.10 trials) did not differ reliably [$F(1,37) = 1.58$, $p > .20$]. (Indeed, most of the difference derives from four subjects in the selective reminding group, each of whom hovered at 19/20 for 3 trials before reaching criterion. Nothing analogous happened in the standard condition.) Using the standard power of .8 as a criterion, this analysis could have detected a difference of 1.74 trials. Thus, trials to acquisition are quite comparable for the two procedures, confirming the finding of Experiment 1.

Figure 2 displays the average number of items correctly recalled on each acquisition trial, computed as in Experiment 1. A comparison of the two groups using number correct on each trial (collapsed over subjects) demonstrates equivalent rates of approach to criterion ($F < 1$). With a power of .8, this analysis could have detected a difference between means of 3.11 items. This nonsignificant finding confirms that of Experiment 1. Also, inspection of Figure 2 shows that any criterion for acquisition would result in the two conditions' being quite equivalent.

How many items actually were presented during acquisition to the subjects in each group? As in Experiment 1, the subjects saw many fewer items over trials in the selective reminding group (mean of 45.15 items) than in the standard group (mean of 102.10 items). This difference was highly significant [$F(1,37) = 103.49$, $p < .001$], replicating the finding in Experiment 1. Thus, acquisition can take much longer for subjects who learn via the standard procedure.

Finally, to examine the microstructure of item recall on acquisition tests, an analysis of subjective organization was undertaken, as in Experiment 1. The degree of clustering increased over trials in both groups in almost identical fashion. The median SO scores across the first seven acquisition trials (for which data were available for both groups) were .16, .16, .26, .26, .26, and .35 in the standard group and .13, .15, .27, .27, .32, and .25 in the selective reminding group. Not only was overall

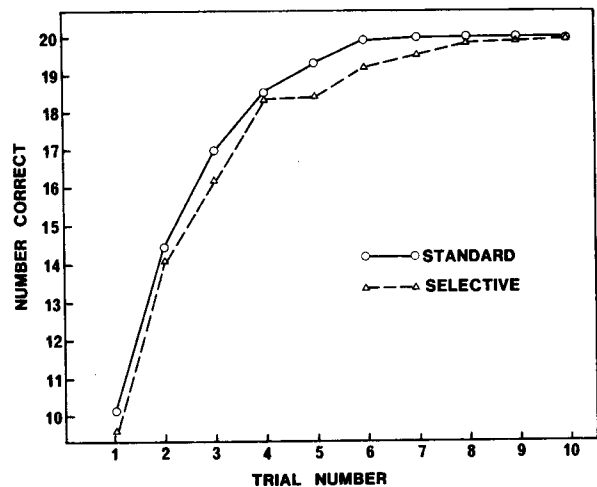


Figure 2. Experiment 2: Number of items correctly recalled on each successive test trial during acquisition as a function of acquisition procedure.

degree of subjective organization virtually identical between groups, but also visual inspection suggested that the same clusters of words tended to be used in both groups. This solidifies the result observed in Experiment 1 and strengthens the conclusion that subjective organization does not differ over the two conditions.

To summarize the acquisition data, both rate of approach to criterion and number of trials to criterion are quite comparable under the two list-learning procedures, as is degree of subjective organization over trials. However, a subject in the selective reminding condition sees many fewer item presentations over trials than does a subject in the standard condition. These findings replicate those of Experiment 1 in suggesting that selective reminding is a more efficient acquisition procedure than is the standard procedure of presenting the entire list on every study trial. Once again, there is no evidence of any drawbacks associated with the selective reminding procedure.

Retention. Like Experiment 1, Experiment 2 examined retention of the list after a 3-week delay. Once again, error rate on the delayed recall test was very low. Of the total of 546 words recalled by all subjects, only 63 were incorrect, yielding an intrusion rate of 11.2% and a repetition rate of 0.4%.³ There were only 14 errors on the 538 words recalled during relearning, 3 extralist intrusions, 2 repetitions, and 9 original-list intrusions. All errors on both tests were split evenly between the two list-learning procedures.

On the delayed retention test, the mean numbers of correctly recalled items were 11.64 for the selective reminding group and 13.15 for the standard group, a difference of 1.5 items. This difference was not significant [$F(1,37) = 1.54$, $p > .20$]. Using a power of .8, this analysis could have detected a difference of 3.57 items. As in Experiment 1, greater variance in the delayed-recall data increased the detectable difference. The important thing is

that the two experiments agree. A list is remembered about equally well regardless of the procedure by which it was learned originally.

Subjective organization can be indexed by pair repetitions from the criterion acquisition test trial on the delayed recall test. The identical median SO scores of .18 for both groups demonstrate that the amount of clustering is similar in retention, just as it was in acquisition. Inspection of the protocols again shows that certain items form pairs quite consistently (e.g., NORWAY-CHICAGO, LEMON-GARLIC), accounting for the bulk of the clustering apparent in both conditions.

In Experiment 2, the relearning test data were no longer subject to a ceiling effect. For the standard group, 8.26 of the 10 original items were recalled correctly, but only 5.15 of the 10 new items. For the selective reminding group, the means were 8.60 for original items and 5.05 for new items. A 2×2 analysis of variance revealed a significant effect of item type (original vs. new) [$F(1,37) = 87.38, p < .001$]. However, the main effect of group and the interaction were both nonsignificant. Thus, it is now legitimate to claim that neither relearning of original items nor learning of new items is affected by which of the two acquisition procedures was used originally.

To summarize the delayed retention data, neither delayed recall nor relearning displayed any differences attributable to whether acquisition was via the standard procedure or the selective reminding procedure. Also, the two groups did not differ in subjective list organization after the retention interval. As was the case in Experiment 1, the retention test results complement the findings for acquisition.

GENERAL DISCUSSION

Virtually all of the published research investigating learning and memory via the free-recall task has used the *standard* acquisition procedure. Here, a list is presented for study in its entirety until some criterion level of recall accuracy is attained. Buschke (1973) introduced *selective reminding* procedure particularly for the purpose of studying organization and clustering in free recall, but this procedure clearly provides an alternative, general acquisition procedure. The present study compared these two ways of learning a list for free recall and, in two experiments, found them to be almost identical in terms of both acquisition and later retention performance. The only observed difference between the two procedures was the number of times items were presented over study trials, a difference that strongly favored selective reminding. In all other respects—trials to criterion, rate of acquisition, error patterns during acquisition and retention, subjective organization during acquisition and retention, accuracy in delayed retention, and success in relearning—the two list-learning procedures were highly similar. To borrow the conclusion offered by Murdock et al. (1974) for their reordering procedure, "The characteristics of

list presentation seem to be surprisingly unimportant in affecting the learning itself" (p. 528).

From a theoretical standpoint, it may seem surprising that learning and retention are not affected by the choice of acquisition procedure. Indeed, three sources of potential differences were considered in the introduction: (1) greater overlearning under the standard procedure, (2) more consistent organization under the standard procedure, and (3) better retrieval information under selective reminding. There is no evidence in support of any of these. If the standard procedure caused greater overlearning by presenting already learned items continuously over trials, this might be expected to affect delayed recall. Specifically, either overall delayed recall or delayed recall of those items learned early in acquisition should be superior under the standard condition. Neither of these results eventuated in either experiment. If the standard procedure led to greater organization because the entire list was seen on each study trial, subjective organization should have developed faster and been more extensive than under the selective reminding procedure. Again, the evidence does not support these predictions. Finally, if selective reminding provided better retrieval information by identifying difficult-to-learn items, acquisition might have been expected to be accelerated in this condition. This was not the case; trials to criterion were equivalent in the two conditions.

Why, then, are acquisition and retention so similar under the two learning regimens? It could be argued that two or more processes trade off in memory so as to produce no observable difference between the two conditions. For example, the retrieval information advantage under selective reminding might offset the overlearning advantage under the standard condition. But such an argument seems very contrived. A better way of understanding the absence of any difference may be to focus on what the subject is doing during acquisition. Because the goal on every test trial is to recall as many list words as possible, the subject continually rehearses the remembered words. Each study trial represents an opportunity to add more words to the remembered set. In the standard condition, the subject must locate the yet-to-be-learned items in the complete set, presumably ignoring already learned items in the process. The selective reminding procedure simply performs this location for the subject, leaving all other processes identical to those in the standard procedure.

If this simple analysis is correct, then there is a clear-cut practical implication of these findings. The selective reminding procedure has the advantage of being faster to administer than the standard procedure because fewer items need be presented on each study trial. Coupled with the finding that selective reminding has no apparent drawbacks compared with the standard procedure, this recommends selective reminding when subjects are to be taught a list for free recall. Unless type of acquisition itself is under study, selective reminding probably is a more efficient technique for list learning.

REFERENCES

- ATKINSON, R. C., & PAULSON, J. A. (1972). An approach to the psychology of instruction. *Psychological Bulletin*, *78*, 49-61.
- BATTIG, W. F. (1965). Procedural problems in paired-associate learning research. *Psychonomic Monograph Supplements*, *1*(No. 1).
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*(3, Pt. 2).
- BOUSFIELD, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229-240.
- BUSCHKE, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior*, *12*, 543-550.
- BUSCHKE, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior*, *15*, 313-324.
- FULD, P. A., & BUSCHKE, H. (1976). Stages of retrieval in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, *15*, 401-410.
- KARUSH, W., & DEAR, R. E. (1966). Optimal stimulus presentation strategy for a stimulus sampling model of learning. *Journal of Mathematical Psychology*, *3*, 19-47.
- KEPPEL, G. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.
- KINTSCH, W. (1970). *Learning, memory, and conceptual processes*. New York: Wiley.
- LORTON, P., JR. (1972). *Computer-based instruction in spelling: An investigation of optimal strategies for presenting instructional material*. Unpublished doctoral dissertation, Stanford University.
- MACLEOD, C. M. (1976). Bilingual episodic memory: Acquisition and forgetting. *Journal of Verbal Learning and Verbal Behavior*, *15*, 347-364.
- MURDOCK, B. B., JR., ANDERSON, R. E., & HO, E. (1974). Effects of presentation order on learning in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior*, *13*, 522-529.
- MURDOCK, B. B., JR., PENNEY, C., & AAMIRY, A. (1970). Interactive presentation on multi-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*, 679-683.
- MURPHY, M. D., & PUFF, C. R. (1982). Free recall: Basic methodology and analyses. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition*. New York: Academic Press.
- NELSON, T. O. (1971). Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 568-576.
- NELSON, T. O., & MACLEOD, C. M. (1974). Fluctuations in recall across successive test trials. *Memory & Cognition*, *2*, 687-690.
- POSTMAN, L. (1962). Repetition and paired-associate learning. *American Journal of Psychology*, *75*, 372-389.
- ROCK, I. (1957). The role of repetition in associative learning. *American Journal of Psychology*, *70*, 186-193.
- TULVING, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*, 344-354.
- TULVING, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175-184.

NOTES

1. Because subjects completed acquisition in varying numbers of trials, there are two ways to plot Figure 1: (1) Drop out subjects as they reach criterion, so that fewer subjects are included as trial number increases, or (2) give subjects credit for perfect performance on postcriterial trials, so that sample size is constant on all trials. I chose the second approach. Although neither is ideal in all respects, they generate roughly the same result here, but the second approach avoids allowing one or two subjects to distort the curve on later trials.
2. Because of the way the program for collecting acquisition data was written, errors and repetitions were not recorded during acquisition. However, the error results for Experiment 1 suggest that this is not a serious shortcoming because these data would have revealed very little.
3. The rate of intrusion is inflated dramatically by two subjects, one in each group, both of whom apparently adopted the strategy of producing precisely 20 words on the delayed test, rather than just those they actually remembered. If just the data of these two subjects are set aside, the intrusion rate drops to 6.6%.

(Manuscript received December 12, 1984;
revision accepted for publication December 27, 1984.)