JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 12, 675-682 (1973)

Test Appropriate Strategies in Retention of Categorized Lists[1]

LARRY L. JACOBY

Iowa State University

Four categorized lists were presented for a single study and test trial. Form of retention test (recognition, cued recall, or free recall) for Lists 1-3 was factorially combined with that of List 4. Learning to learn was evident only for cued recall and improvement in that condition was primarily due to an increase in the number of items per category recalled. Effects in Test 4 performance provided evidence that study strategy depended on the form of test anticipated. Subjects anticipating a cued recall test apparently spent less time studying category names and more time on the study of category instances than did subjects preparing for free recall. Implications of test-appropriate study strategies for theories of memory are considered.

Prior to a classroom test, students often request information concerning the form of test that is to be given. One gets the impression that they plan to spend more time integrating the material if they are told that they will receive an essay rather than a short-answer or recognition test. Several memory theorists have shared students' intuitions with regard to differences in test requirements. Kintsch (1970) has suggested that the primary influence of organization is on retrieval of presented material; a process that is said to be important for recall but involved in only a trivial manner in recognition. Underwood (1972) has taken a similar position by suggesting that the attributes used for recognition might not include associative attributes that are essential for recall. Further, Underwood stated that subjects may be able to influence the memory composition of an item if they can anticipate the form of the test. That is, subjects may learn to encode information in a form that allows maximal test performance.

The memory representation of an item can be visualized as being hierarchical with information at higher levels being abstracted from presentation of list items. The accessibility of higher-level information might be necessary to allow retrieval of information at lower levels. For example, it might be necessary to retrieve category names prior to the retrieval of presented category instances. Thus, preparation for a test may necessitate the study of information at several different levels of abstraction. If subjects can anticipate the form of the impending test, they are then free to focus their study on required information that the test will not provide. The result would be a study strategy that is optimal for the particular test form but might be quite inefficient for tests of other types.

Most theorizing has centered around differences in free recall and recognition tests. However, a continuum with regard to the number of search or retrieval cues provided by a test can be visualized (Shriffrin, 1970). Recognition and free recall would serve as endpoints on this continuum with cued recall falling between the two. A minimal number of retrieval cues is provided by a free recall test. Free recall of a categorized list requires that a subject be prepared to retrieve both category names and category instances. The necessity of retrieving categories can be largely removed by employing a cued recall test that provides category names. If a cued recall test is anticipated, subjects may be able to spend less time studying category names and additional time

[1] The author expresses appreciation to Eric Martell for his assistance in the collection and scoring of data.

on the study of category instances. The number of retrieval cues provided by a recognition test is near maximal since a test item may be used as a cue for retrieval of its own memory representation. As a consequence, the optimal study strategy for a recognition test might be expected to differ substantially from that for a free recall test.

The logic of the present experiment was similar to that of investigations of learning to learn (e.g., Postman, 1969). The method of test (free recall, cued recall, or recognition) was held constant across three lists. This consistency of test should lead subjects to expect the same form of test following a fourth study list. The form of fourth test actually given was combined factorially with that of the first three lists. For example, free recall of the fourth list followed free recall, cued recall, or recognition of the first three lists. Differences in performance on the common fourth test can be attributed to strategies developed across prior lists. Highest performance when test form had been held constant across all lists would indicate that study strategies specific to the particular form of test had been developed.

METHOD

Materials

The 14 most frequently reported instances were selected from each of 32 categories listed in the Battig and Montague (1969) norms. Words that held an odd-numbered frequency rank in the norms were employed as study items while those holding an even-numbered rank served as new distractor items for the recognition tests. Four 56-item study lists were formed with each list containing seven instances each of eight different categories; instances of a category were blocked during study presentation.

Retention of each list was assessed by means of either a free recall, cued recall, or recognition test. A separate test booklet was prepared for each of the test forms. The first page of each booklet was blank with the exception of a statement that informed subjects that they were not to turn that page until instructed to do so. Instructions for the test that was to be given were presented on the second page of each test booklet. Free recall instructions informed subjects that they were to write on the following page of the test booklet all of the words they could remember from the list just presented. Instructions in cued recall test booklets informed subjects that each of the following pages would contain category names. They were instructed to write beneath each category name words from that category that had occurred in the study list just presented. Each of four pages in the test booklet contained two category names; categories were randomly assigned to test position. Recognition instructions informed subjects that each of the following pages would contain a column of words and that they were to circle words that had been presented in the study list. Each of four pages contained a single column of 28 words. Study and distractor items were from the same categories and equal in number; items were randomly assigned to test position with the restriction that two instances of the same category could not occur in adjacent positions.

Design and Subjects

Form of retention test (free recall, cued recall, or recognition) was held constant across the first three lists studied and factorially combined with form of test given on the fourth list (free recall, cued recall, or recognition). The resulting design was a 3×3 factorial; both factors were manipulated between subjects. Four replications of this basic design were formed by rotating lists through study order so that each list served equally often as the first, second, third, and fourth list studied.

The subjects were 144 volunteers enrolled in an introductory psychology course at Iowa State University; 16 subjects were assigned to each of the experimental conditions. Since subjects were tested in small groups ranging in size from two to five, it was necessary to randomly assign groups rather than individual subjects to experimental conditions.

Procedure

Study lists were videotaped and presented for a single study trial at a rate of 2 sec per item. Subjects were informed that they would see and be tested on four lists, and that each list would contain 56 words arranged in categories. At the end of each list, the phrase "Begin Test" was presented. This was the signal to open the test booklet, read instructions, and then proceed to complete the test. The subjects had no means of anticipating the form of test prior to reading test booklet instructions. At the end of the 5-min. test period, the experimenter instructed subjects to close their test booklet and place it at the bottom of the pile of test booklets in front of them. This procedure was repeated until all four lists had been presented and tested.

Results and Discussion

Tests 1–3

Performance on Tests 1–3 was analyzed separately for the three types of retention test. All analyses included form of fourth test as a factor. This factor did not approach significance in any of the analyses. Thus, there was no evidence of differences among fourth-test conditions prior to differential experimental treatment.

Free and cued recall. Mean correct responses and intrusion errors from Tests 1–3 are presented in Table 1. Additional measures included in Table 1 were employed to separate influences on category recall from those on recall of category instances. Category recall was defined as the number of categories from which at least one word was recalled. The items per category measure (IPC) was defined as the ratio of the number of words recalled to the number of categories recalled. The definition of measures was identical for the free and cued recall conditions. Both the category and IPC measures have been used by other investigators (e.g., Tulving & Pearlstone, 1966).

TABLE 1

Free (FR) and Cued (CR) Recall Statistics from Tests 1–3

Statistic	Test number 1	2	3
Correct responses			
FR	20.9	22.1	21.5
CR	29.2	34.6	36.2
Category intrusions			
FR	1.1	1.5	1.7
CR	2.1	2.0	1.7
Categories recalled			
FR	5.7	5.5	5.4
CR	7.8	7.9	8.0
Items per category			
FR	3.6	4.1	4.0
CR	3.7	4.4	4.6

The number of correct free recall responses did not change significantly across successive tests. In contrast, there was a substantial increase in cued-recall correct responses, $F(2, 94) = 42.40$, $p < .001$. The frequency of intrusion errors did not vary significantly as a function of test number in either the free or cued recall condition.

Category recall was near perfect on all cued recall tests. The number of categories recalled was lower in the free recall condition and showed a slight, nonsignificant decline across successive tests. Mean IPC recall from Tests 1–3 is shown in the bottom rows of Table 1. An analysis of these data revealed that IPC recall increased across tests, $F(2, 188) = 36.52$, $p < .001$, and was higher for cued than for free recall, $F(1, 94) = 7.99, p < .01$. The interaction of test form and number was also significant, $F(2, 188) = 6.40$, $p < .01$. Additional analyses revealed that the increase in IPC recall was significant in both the free, $F(2, 94) = 8.51$, and the cued, $F(2, 94) = 34.44$, recall conditions, $p < .001$ for both.

Recognition. Mean correct recognitions (hits) and errors (false alarms) from Tests 1–3 are shown in Table 2. Both the frequency of hits and false alarms remained quite stable across tests. A difference score was computed for each subject by subtracting the number of false alarms from hits; the signal detection model was employed to obtain d' as a second measure of recognition. The effect of test number did not approach significance, $F < 1$, in the analysis of either measure.

TABLE 2

Recognition Statistics from Tests 1–3

Statistic	Test number 1	2	3
Hits	42.29	42.94	43.19
False alarms	7.85	8.06	8.06

The present recall results can be compared with those of Tulving and Pearlstone (1966). Tulving and Pearlstone exposed subjects to a single study and test trial of a categorized list and found that cued recall produced higher performance than did free recall. Further analyses revealed that the cued recall advantage was totally due to a larger number of categories recalled; IPC recall did not differ for the two types of tests. The first test trial results of the present investigation were in agreement with the results reported by Tulving and Pearlstone. As shown in Table 1, the cued condition produced a higher level of word and category recall than did the free recall condition. On the first test, mean IPC recall was nearly identical in the two test conditions so that the word recall advantage of the cued condition can be totally attributed to differences in category recall on later tests, however, cued recall held an advantage in both IPC and category recall. There were apparently both storage and retrieval differences between the free and cued recall conditions on the later tests. When a cued-recall test was anticipated, subjects were able to spend additional time studying category instances.

Variation in number of correct responses across Tests 1–3 might be demanded as evidence of strategy development. If so, there is clear evidence of strategy development only for the cued recall condition. The free recall condition showed learning to learn in IPC recall but a corresponding decrease in category recall; the result was that the total number of correct free recall responses remained relatively stable across tests. There was no evidence of learning to learn across successive recognition tests.

Test 4

The influence of preceding test form on fourth test performance provides an additional means of assessing strategy development. If strategies appropriate to a particular test have been developed, fourth test performance should be highest when that test is of the same form as the three preceding tests. Differences among conditions in fourth test category and IPC recall provide information concerning the nature of strategies that have been developed.

Mean fourth test correct responses and errors are presented in Table 3 for each combination of test conditions. Category and IPC means are also included in Table 3 for conditions engaging in free or cued recall as a fourth test. As was the case for Tests 1–3, category recall was defined as the number of categories from which at least one word was recalled while IPC recall was defined as the ratio of the number of words recalled to the number of categories recalled.

TABLE 3

FREE RECALL (FR), CUED RECALL (CR), AND RECOGNITION STATISTICS FROM TEST 4

Statistic for test 4	Preceding test: FR	CR	Recognition
Correct responses			
FR	20.8	20.1	20.9
CR	29.6	36.8	27.9
Recognition	41.9	47.9	43.5
Category intrusions			
FR	.8	1.3	1.0
CR	2.7	3.0	3.2
Recognition[a]	10.0	9.2	9.0
Categories entered			
FR	5.2	4.4	5.1
CR	7.8	8.0	7.9
Words per category			
FR	4.0	4.6	4.1
CR	3.8	4.6	3.5

[a] These are number of false alarms.

Free and cued recall. With free recall as a fourth test, the effect of preceding test form did not approach significance in either the analysis of correct responses or that of intrusion errors, $F < 1$. The frequency of free recall correct responses and errors on the fourth test were nearly identical to those on the first free recall test. When the fourth test was one of cued recall, the influence of preceding test form on correct responses was significant, $F(2, 45) = 9.45$, $p < .01$. Correct cued recall responses were more frequent after cued recall on Tests 1–3 than after either free recall or recognition. The number of cued recall intrusions did not differ significantly among conditions.

Category recall was near perfect for all conditions when recall was cued on the fourth test. Free recall as a fourth test produced a lower level of category recall, and differences among preceding test conditions. Although the effect was not significant, $F(2, 45) = 1.27$, $p > .10$, free recall of categories was numerically lower after cued recall than after either recognition or free recall.

Differences among preceding test conditions should be most pronounced for categories represented in the early portion of a study list. Owing to their recency, recall of categories represented near the end of a list might not reflect differences in study strategy. A further analysis related free recall of categories to the study list position (input block) of category instances. Category recall as a function of input block is presented in Figure 1 for conditions that engaged in either free or cued recall on Tests 1–3. The category recall curve from the first free recall test is also presented in Figure 1 for purposes of comparison. A plot of number of words recalled from each input block (not conditionalized on category recall) was nearly identical to that shown in Figure 1 so that the curves are descriptive of word as well as category recall probability.

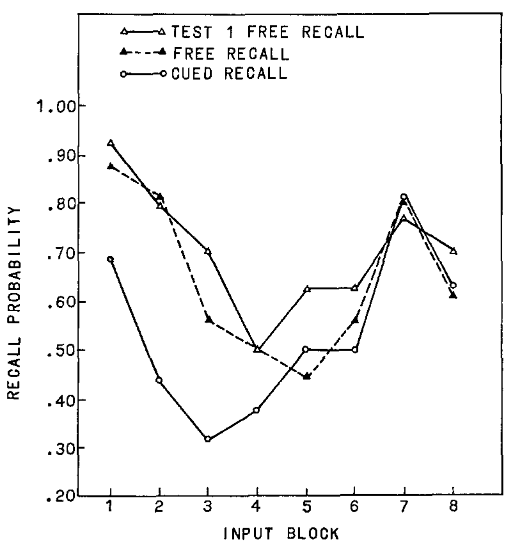

FIG. 1. Category recall probablity on free recall test 4 as a function of prior test form and input block.

Fewer categories were free recalled from Blocks 1–4 after cued than after free recall of the first three lists, $F(1, 45) = 6.03$, $p < .025$; recall from later blocks was nearly identical for the two conditions. Comparisons with the curve from the first free recall test suggest that changes in study strategy developed in the cued but not in the free recall condition. Subjects anticipating a cued recall test apparently spent less time on the study of categories.

The influence of study strategy on the effect of position within a category was the topic of an additional analysis. Items free recalled on the fourth test were classified by position within an input block for conditions that had engaged in either free or cued recall on Tests 1–3. The analysis was conducted only on items that had occurred within Blocks 2–7; items from Block 1 and 8 were eliminated from the analysis due to the possibility of primacy and recency effects.

Results of the above analysis revealed a marginally significant, $F(6, 180) = 2.10$, $p < .05$, main effect of position within a category. The first-presented instances of a category were recalled with a higher probability than were later-presented instance (.37, .36, .32, .35, .31, .26, and .29). The interaction of preceding test condition and position within a category did not approach significance, $F(6, 180) = 1.35$, $p > .10$. Thus, preparation for cued recall influenced the recall of categories but did not alter the effect of position within a category.

Free and cued recall IPC means from Test 4 are presented in the last rows of Table 3. With either type of fourth test, IPC recall was highest when Tests 1–3 were cued, $F(2, 90) = 12.16$, $p < .001$. Recognition and free-recall conditions tended to produce higher IPC recall when the fourth test was free rather than cued; preceding cued recall produced nearly identical IPC performance on free and cued recall tests. Improvement in cued IPC recall on the fourth as compared to the first test was evident only for the condition that engaged in cued recall on all tests. All conditions produced higher IPC recall on the fourth free recall test than was evident on the first test of that type.

Recognition. Mean recognition hits and false alarms are also presented in Table 3. The number of hits after cued recall was significantly larger than that after either recognition, $F(1, 45) = 4.97$, $p < .05$, or free recall, $F(1, 45) = 8.59$, $p < .01$; the number of false alarms did not differ significantly among conditions. Only the difference between preceding free and cued recall was significant when recognition was corrected for guessing by subtracting false alarms from hits, $F(1, 45) = 4.36$, $p < .05$, or by employing d' scores, $F(1, 45) = 4.76$, $p < .05$. Further analyses failed to reveal any significant effects or interactions involving input block in either the probability of a hit or false alarm.

Discussion

The existence of test-appropriate study strategies is of considerable theoretical importance. Several theorists (e.g., Postman, 1963) have attempted to account for all memory effects by postulating variation along a single hypothetical dimension. For example, a "strength" theory postulates that test performance is a function of the memory strength of to-be-remembered items (Postman, 1963). The memory requirements of recognition and recall tests have been said to differ only with regard to the degree of memory strength necessary to allow a correct response. Presumably, cued recall would require a memory-strength intermediate to those required for free recall and recognition. A strength theory would predict that, within limits imposed by ceiling effects, any variation capable of increasing performance level on one type of test would also enhance performance on tests of all other types. The existence of test-appropriate study strategies is incompatible with strength theory or any other unidimensional theory of memory.

In an influential paper, Tulving and Pearlstone (1966) made a distinction between the availability and the accessibility of items in memory. The superiority of cued over free recall was given as evidence that information sufficient to recall additional items was available in memory but not accessible during a free recall test. Given the distinction between availability and accessibility, a major problem is determining what factors influence accessibility. If category names are used as retrieval cues for category instances in free recall, what factors influence the retrievability of category names?

The answer to the above question might be that category names are abstracted and studied independently of category instances. That is, relationships among items might be abstracted and studied separately so as to increase their retention as retrieval cues for presented items. The memory representation of an item can be viewed as being hierarchical with information at higher levels being abstracted from presented items. Retrieval may be constrained by this hierarchy so that higher-level information must be accessible prior to the retrieval of lower-level information. Inaccessibility of information in memory would then imply that other information, at a higher level in the hierarchy, was not available at the time of test. Study might be distributed among activities concerned with construction of the hierarchy and other activities designed to allow retention of points in the hierarchy once they have been constructed. For example, additional study of a category name after it has been abstracted might be necessary to allow its later retention. Retention of information at any level in the hierarchy would then be relatively independent of that at other levels, and a direct function of the quantity of study it had received.

The present result revealed that retention of information at different levels of abstraction was influenced by study strategy. Subjects anticipating a cued test were able to free recall fewer categories but more instances of each recalled category than were subjects that anticipated a free recall test. Within the framework presented above, these results can be interpreted as evidence that subjects preparing for a cued test spent more time studying category instances and less time studying category names. Cued recall and recognition results from the fourth test support the claim that less time was spent studying category instances when a free recall test was anticipated; both fourth test recognition and cued recall performance were lower when subjects anticipated free rather than cued recall. Thus, retention of category names and retention of category instances were found to be relatively independent, and influenced by study strategy.

It was earlier stated that a cued recall test eliminates the necessity of retrieving categories. The problem one faces with a statement of this type is identical to that encountered by theories of recognition memory. Recognition apparently depends on the similarity of the encoded version of a test item and its study counterpart (Jacoby & Hendricks, in press; Tulving & Thomson, 1971). Correspondingly, the cue effectiveness of a category name would be expected to depend on the similarity of its encoding to the category representation stored during study. Martin (1972) has voiced concern with an identical problem in paired-associate learning by noting that the encoding of a stimulus might vary between presentations. The problem is even more pronounced in the cued recall case when the category name is not presented during study but must be abstracted from category instances. The category name may be used as a cue for retrieval of a representation encoded during study that is only similar, not identical to the category name. One advantage of repeated testing might be that subjects learn to increase the similarity of representations encoded during study to the cues that will be offered by the test.

Study strategies that were specific to either free recall or recognition were apparently not developed in the present investigation. When subjects are unable to anticipate the form of test, they might behave in a conservative fashion and prepare for free recall. A preexperimental strategy appropriate for free recall might then have been employed from the outset; both category names and instances may have been studied beginning with the first list presented. Subjects receiving free recall tests would have no reason to modify their preexperimental strategy. It is surprising, however, that the initial strategy was not modified appreciably by subjects receiving recognition tests. A potential explanation is that the recognition test did not provide unambiguous category information (Jacoby, 1972). There was no assurance that all recognition test items were from a category presented during study, and there was also the danger of categorizing a test item differently from the way in which it had been categorized during study. Thus, category information may have been studied and used to aid recognition performance. Additional research is needed to clarify differences in memory requirements of free recall and recognition tests.

References

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

JACOBY, L. L. Effects of organization on recognition memory. *Journal of Experimental Psychology*, 1972, **92**, 325–331.

JACOBY, L. L., & HENDRICKS, R. L. Recognition effects of study organization and test context. *Journal of Experimental Psychology*, in press.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

MARTIN, E. Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.

POSTMAN, L. One-trial learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.

POSTMAN, L. Experimental analysis of learning to learn. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 3. New York: Academic Press, 1969.

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*, New York: Academic Press, 1970.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

TULVING, E., & THOMSON, D. W. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–125.

UNDERWOOD, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.

(Received May 18, 1973)